DOCUMENT RESUME

ED 404 175 SE 059 717

AUTHOR	Muthen, Bengt O.; And Others
TITLE	Multidimensional Description of Subgroup Differences in Mathematics Achievement Data from the 1992 National Assessment of Educational Progress. Draft.
INSTITUTION	National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NAEP-92-123093
PUB DATE	Jan 94
NOTE	90p.; Some tables are out of order and some tables are missing data.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Cultural Differences; Elementary Secondary Education; *Mathematics Achievement; Mathematics Education; *Racial Differences; *Sex Differences; *Test Construction
IDENTIFIERS	*National Assessment of Educational Progress

ABSTRACT

This report investigates the dimensionality of the 1992 National Assessment of Educational Progress (NAEP) mathematics test in the context of subgroup differences. The analysis approach of this study utilized key grouping variables of the NAEP reports (e.g., gender, ethnicity), but had the advantage that subgroup comparisons were not only done in a univariate manner using one grouping variable at a time, but were done using the set of grouping variables jointly. The data supports a multidimensional model with dimensions corresponding to both content-specific and format-specific factors. The multidimensional latent variable modeling suggests a new way of reporting results with respect to math performance in specific content areas. For content-specific performance, the subscores were related to overall performance, considering content-specific scores conditional on overall scores. For a given overall score a subgroup difference was considered with respect to a certain content area. This conditional approach may be of value for revealing differences in opportunity to learn or differences in curricular emphases. (Author/JRH)

SE

CRESST

National Center for Research
on Evaluation, Standards,
and Student Testing

ED 404 175

Multidimensional Description of Subgroup Differences in Mathematics Achievement Data from the 1992 National Assessment of Educational Progress

- UCLA Center for the Study of Evaluation

in collaboration with:

- University of Colorado
- NORC, University of Chicago
- LRDC, University of Pittsburgh
- University of California, Santa Barbara
- University of Southern California
- The RAND Corporation

TM025717

January 1994
Draft

Multidimensional Description of Subgroup Differences in Mathematics Achievement Data from the 1992 National Assessment of Educational Progress

Bengt O. Muthén

Siek-Toon Khoo

Ginger Nelson Goff

Graduate School of Education

UCLA

Los Angeles, CA 90024-1521

This research was supported by the National Center for Education Statistics. I am thankful for the research assistance of Li-Chiao Huang, Guanghan Liu, and Todd Franke and comments from Leigh Burstein, Irene Grohar, and Linda Winfield.

NAEP92Paper123093

Abstract

This report investigates the dimensionality of the 1992 NAEP mathematics test in the context of subgroup differences. A multidimensional model is supported by these data with dimensions corresponding to both content-specific and format-specific factors. The analysis approach of this paper utilizes key grouping variables of the NAEP reports (e.g., gender, ethnicity), but has the advantage that subgroup comparisons are not only done in a univariate manner using one grouping variable at a time, but using the set of grouping variables jointly. This is carried out within a structural model with latent variables, which relates the information on the test items to background information via a set of factors. It is found that the different factors relate differently to the background variables. The multidimensional latent variable modeling also suggests a new way of reporting results with respect to math performance in specific content areas. For content-specific performance, the subscores are related to overall performance, considering content-specific scores conditional on overall scores. For a given overall score a subgroup difference is considered with respect to a certain content area. This conditional approach may be of value for revealing differences in opportunity to learn or differences in curricular emphases. Conditional differences may be viewed as "unrealized potential" for performance in a specific content area.

Introduction

This report examines mathematics achievement data from the National Assessment of Educational Progress (NAEP). NAEP is a regularly administered, Congressionally mandated assessment program for the nation and the states. NAEP test results for grades 4, 8, and 12 are reported for various subgroups of the U.S. school population. The most recent mathematics report, "NAEP 1992 Mathematics Report Card for the Nation and the States" (Mullis, Dossey, Owen, Phillips, 1993), includes overall mathematics proficiencies for subgroups based on region, gender, ethnicity, type of community, parents' highest level of education, and type of school. Proficiencies for the entire group are also reported for the specific content areas of numbers and operations; measurement; geometry; data analysis, statistics, and probability; and algebra and functions. Content-specific subgroup comparisons are given in the NAEP Data Almanacs.

The aim of this report is to investigate the dimensionality of the mathematics test. This test consists of a large number of items distributed over a number of test forms to which students are randomly assigned. In analyzing 1990 NAEP math data, it was suggested that the math items are essentially unidimensional with respect to content areas with the possible exception of geometry in grade 8 (Rock, 1991). Support for unidimensionality is usually based on finding correlations close to unity among factors representing various aspects of the items. Rock's analysis of content areas showed correlations in the range 0.86 - 0.95 for grades four, eight, and twelve. Unidimensionality was also indicated in analyses considering item format (Carlson & Jirele, 1993). Using the 1992 data a more detailed analysis with respect to item format was given in Mazzeo, Yamamoto, and Kulick (1993). The 1992 test included both short constructed-response items and extended constructed-response items in addition to the traditional item format of multiple-choice items. The Mazzeo et al. analysis found an important deviation from unidimensionality only for extended constructed-response items. In 1992, however, extended constructed-response items made up less than 4% of the total number of items for grades 4, 8, and 12.

As mentioned above, NAEP reports subgroup differences with respect to overall math performance, whereas content-specific performance is typically not reported for subgroups. Given the indications of unidimensionality, one may in fact ask if content-specific reporting is at all necessary, or if the overall reporting is sufficient. The idea of simplified reporting has been discussed among ETS researchers. For example, in analyzing 1990 NAEP math data Rock (1991) concluded that "there seems to be little discriminant validity here. In conclusion, it would seem that we are doing little damage in using a composite score."

In our view, entertaining the notion of unidimensionality, although useful for simplified reporting, may leave interesting features of the data unexplored. As shown in the appendix, it is not hard to settle for unidimensionality unless a special effort is made to find meaningful additional dimensions. This paper argues that the need for a multidimensional representation of the data is difficult to judge based on the conventional approach reported above of estimating correlations in multifactorial models. This paper goes beyond the conventional approach in two respects. First, it uses a latent variable model that is more sensitive to capturing deviations from unidimensionality.

Using this model, it is shown that there are several additional dimensions that are statistically significant. Second, to evaluate the practical significance of adding these further dimensions, the same subgroups that the NAEP compares are also compared using the multidimensional model.

NAEP's estimation of subgroup differences is based on a statistically-complex procedure where proficiencies are estimated based not only on student performance, but also on background variables ("conditioning variables") including those used for subgroups in the reports. The methodology of this paper utilizes the key grouping variables of the NAEP reports (e.g., gender, ethnicity), but has the advantage that subgroup comparisons are done not only in a univariate manner using one grouping variable at a time, but using the set of grouping variables jointly. This is carried out within a structural model with latent variables, which relates the information on the test items to background information. In this way, the structural model is similar to the framework used by NAEP to produce proficiencies for the subgroups. The results are not, however, arrived at by first estimating proficiencies using conditioning variables. In this way, our methodology has the further benefit of providing a validation of the NAEP procedure.

The multidimensional latent variable modeling used here also suggests a new way of reporting results with respect to math performance in specific content areas. For content-specific performance, we propose relating the subscores to overall performance, considering content-specific scores conditional on overall scores. For a given overall score we ask what the subgroup difference is with respect to a certain content area. The results may show that two individuals with the same overall score but belonging to different subgroups are expected to perform quite differently in a particular content area. This conditional approach gives a sharper focus in the reporting. It may be of value for revealing differences in opportunity to learn or differences in curricular emphases. Conditional differences may be viewed as "unrealized potential" for performance in the specific content area.

Method

Samples

Mathematics data from the 1992 NAEP main assessment are used (the "Main Focused-BIB Assessment"). NAEP is a multistage probability sample with three stages of selection: primary sampling units (PSU's) defined by geographical areas, schools within PSU's, and students within schools. In the 1992 NAEP main assessment 26 different test forms were used, each taken by almost 400 students in each of grades 4, 8, and 12, resulting in test results for almost 10,000 students per grade. The analyses in this paper will focus on grade 8 and grade 12. Given missing data on some of the background variables used in the present analyses, the sample sizes are 8,963 for grade 8 and 8,705 for grade 12, corresponding to missing data rates of 13% for grade 8, and 8% for grade 12.

Variables

The 1992 NAEP main assessment considered test items from the five content areas:

(1) Numbers and Operations (whole numbers. fractions. decimals. integers. ratios. proportions, percents, etc.).

(2) Measurement (describing real-world objects using metric, customary, and non-standard units).

(3) Geometry (geometric figures and relationships in one, two and three dimensions).

(4) Data Analysis, Statistics, and Probability (data representation and interpretation).

(5) Algebra and Functions (algebra. elementary functions. trigonometry, discrete mathematics).

There are three formats used for the 1992 math items: conventional multiple-choice items (binary scored), short constructed-response items (binary scored), and extended constructed-response items. The mix of content and format for the test items of each grade is shown in Table 1. It is seen that the grade 8 test is dominated by Number and Operations items, whereas the grade 12 test has as many Algebra items. About one third of the items are short constructed-response items, whereas less than 4% of the items are of the extended constructed-response format.

Insert Table 1

NAEP results are presented as test scores for each of the five content areas and an overall composite score which is a weighted sum of the five content areas. The determination of the weights is based on what is thought important for students to know at a certain grade level. For grade 4, the weights are (using the order of the five content areas given above): 45, 20, 10, 10, 10. For grade 8 they are: 30, 15, 20, 15, 20. For grade 12 they are: 25, 15, 20, 15, 25. It is seen that Numbers & Operations obtains diminishing weight over grades, whereas Geometry and Algebra obtain increasing weights. The weights for grades 8 and 12 correspond roughly to the item content mix shown in Table 1.

NAEP uses a balanced incomplete block ("Focused-BIB") design to distribute the test items across the test forms. There are 13 blocks of items. Each of the 26 test forms ("booklets") consists of three blocks, each block appears in six booklets, and each block appears once with every other block. Tables 2 and 3 show this design for the twelfth and eighth grade tests, also showing how many students took each block in the samples of students used in the present analyses. As is seen from Table 2, this paper uses each block of items to create a set of testlets. A testlet is a sum of binary scored items, where omits are treated as incorrect. The testlets are specific to content area and item format. The column labelled "Format" shows whether a testlet consists of multiple-choice items (M) or short constructed-response items (C). The column labelled "Content" uses the content area numbering given above. As mentioned above, there were very few extended constructed-response items in mathematics. Dimensionality assessment of such few items would not be meaningful given our aggregation of items into testlets and extended constructed-response items are therefore excluded in the present analyses.

Insert Tables 2 and 3

The use of testlets may be critized as drawing on arbitrary item groupings. This is not an important issue here. Given the fact that each testlet is specific to block, content, and format, it generally consists of only 2-3 items, i.e., all items of a certain content and format within a certain block. In this way, there is most often only one way to aggregate the items. A few blocks, however, afford the creation of more than one testlet per content and format and are labelled a, b, c, ... (see e.g., testlets 2-5). Items which share the same stem are always put into the same testlet.

Tables 2 and 3 also show the degree to which the content areas and item formats are covered by the testlets and the 26 independent samples of students. For example, in Table 2 grade 12 constructed-response (C) type algebra (content area 5) is represented by three testlets in booklet 4 and is available for 354 students in this booklet. It is seen that each testlet appears in six booklets so that for example the algebra testlet 48 in grade 12 has data for a total of 2,051 students. Generally speaking, the content- and format-mix of the testlets is similar to that of the NAEP test items shown in Table 1. Exceptions are Measurement in constructed-response format for grade 12 and Algebra in constructed-response format for grade 8 where the items were spread over too many blocks to be represented by testlets. Factors corresponding to these two types of items can therefore not be identified in the present analyses. Tables 2 and 3 will be further referred to below in connection with the description of the modeling.

The achievement variables will be related to a set of background variables shown in Table 4. This set corresponds to the major subgroups used in NAEP reporting. It is also a key set of variables used in the conditioning procedure used in NAEP's estimation of proficiencies in terms of the amount of latent variable variance explained in the conditioning.

Insert Table 4

Analyses

Multidimensional latent variable modeling

We consider a latent variable model for the set of observed variables corresponding to the testlets. A unidimensional model states that a single continuous latent variable accounts for the associations among these variables. In our analyses, we will expand on this model and allow a specific dimension corresponding to each of the five content areas and each of the two formats. We will call this model a GS model (general-factor, specific-factor model). The model is a version of the classic "bi-factor" model used in Holzinger and Swineford (1939). In this way, the variance of a variable is accounted for by up to three different types of systematic sources of variation. The three sources are taken to be orthogonal as in conventional variance component estimation. The first dimension is a general factor representing the general skill required for solving these types of mathematics problems and may be seen as corresponding conceptually to the "overall" math score in NAEP reports. The GS model describes specific factors as residual testlet covariance given the general factor. Deviations from unidimensionality can be described in terms of the variance component for the specific factors relative to the sum of variance components for the general and specific factors. For each variable the model adds a random error component to the systematic components in order to capture measurement error. Given that the testlets are computed from a small number of items, this portion of the observed variable variance is relatively large. Because the unreliability is accounted for, however, this does not cause problems. This error source of variation is a direct function of how testlets were created and is uninteresting in the context of our investigation. Discussions of relative size of variance components for systematic sources will refer to the reliable portion of a variable's variance. The appendix gives a simple example of a GS model and presents some general formulas related to it. In our analyses, the general factor loadings will be allowed to be free, while for simplicity the specific factor loadings are fixed at unity.

Three features of the GS model should be noted. First, ignoring measurement error, the model implies highly correlated content-specific scores when the specific factor variance components are relatively small. In order to compare these results with the content-factor analysis of 1990 NAEP math data by Rock (1991) as well as the correlations among the five 1992 NAEP content scores, it is of interest to also present the correlations among the five content areas as deduced from the estimated model. As discussed in the appendix, these are computed as the correlations among the reliable part of the content variation, purging the observations of measurement error. The correlations can be very high even for sizable specific-factor variance components.

Second, the GS model emphasizes that the content-specific scores contain both general factor variation and specific factor variation (cf. Schmid & Leiman, 1957). If the GS model is not used, but subgroup differences are considered with respect to content-specific observed scores, differences in the underlying dimensions may be obscured. Subgroups may differ in different ways with respect to the different dimensions of variation. For example, one subgroup may have a slightly higher general factor mean than another subgroup, but a much lower specific factor mean. Given that the general factor dominates the variation in the observed scores, the observed score mean difference may turn out to be zero, concealing the large specific-factor difference.

Third, the GS model lends itself to viewing observed scores graphically, separating the general factor mean differences from specific factor mean differences. The idea is to give information corresponding to that of differential item functioning ("item bias"): for a given general "trait" value on the horizontal axis, the vertical axis shows subgroup differences for a specific content area. In line with regression, a conditional expectation function may be plotted for a testlet score, or its reliable part, given the general factor. When the specific factor is orthogonal to the general factor it may be seen as a residual. This residual has different expectation in different subgroups. When the specific factor is correlated with the general factor as in the full model described in the next section, the mean of the specific factor conditional on the general factor is a function of the general factor. Assuming a low specific-factor, general-factor correlation and a low specific-factor to general-factor variance ratio, the variation in this mean across general factor values is, however, likely to be small (e.g., if a bivariate normal distribution is assumed for the general and specific factor). In this way, considering the conditional expectation function for two subgroups, the same slope (or approximately the same slope) but different intercepts are obtained. The intercept difference is of great substantive interest because it shows how differently two individuals with the same overall score but belonging to different subgroups are expected to perform in a particular content or format area. Because the general factor score represents general math skills needed to do well on the overall test, such differences may represent "unrealized potential" (UP) due to lack of opportunity-to-learn. Figure 1 shows this idea graphically for two groups labelled A and B, where group B shows a large UP value relative to the general factor (or overall) difference.

Insert Figure 1

The NAEP data structure provides an important complication in the modeling. This complication is shown in Tables 2 and 3 given above. Each booklet corresponds to an independent sample of students, so that there are 26 independent groups of observations. While there is a total of 49 distinct observed variables (testlets) in grade 12 and 51 in grade 8. for any given group of students only a few of these variables are observed. In this way, the data shows an intricate missing data pattern. Theory for structural equation modeling with missing data patterns of this type has been discussed in Muthén, Kaplan, and Hollis (1987). The solution is a multiple-group analysis where the 26 groups of students are analyzed jointly. Because each observed variable occurs in six of the groups, equalities of parameters involving common variables are applied across groups. Given that the GS model detects specific factors as residual testlet covariance given the general factor, the modeling is dependent on having at least two, and preferable more, testlets per content-and format-specific factor. To have a large enough sample to support stable estimation of specific factors this testlet requirement should hold for at least two booklets. Tables 2 and 3 show that these minimum requirements are fulfilled (for multiple-choice testlets there is always more than two such testlets).

With five content areas and two item formats, ten specific factors can in principle be included in the GS model. To better define the general factor, however, the content area of Numbers & Operations in multiple-choice format will not be represented by a specific factor. These types of items represent central math topics tested in a conventional way. In this way, the general factor is the only factor that influences such testlets and the general factor is therefore defined in terms of performance on these traditional types of items. Alternative specifications which include a specific factor for these types of items show that the results are not sensitive to this choice of "rotation" of the general factor.

A Structural Model for Relating Achievement to Background (MIMIC Modeling)

The multidimensional latent variable model described above will be incorporated in a structural equation model which relates the factors to the set of background variables. This type of analysis is often referred to as MIMIC (multiple-indicators, multiple-causes) modeling in structural equation language. For applications to the study of group differences, see e.g., Muthén (1989). The multidimensional model for the achievement variables provides the measurement part of the structural model. In this part, the estimates of key interest are the percentages of the reliable variance in the observed variables that is due to the specific factors. As mentioned above, these values will be interpreted as the amount of deviation from unidimensionality. The linear regression equations relating the factors to the background variables provide a way to describe mean differences in the factors with respect to the groupings represented by the background variables in a way analogous to dummy variable regression. The MIMIC model is shown in path diagram form in Figure 2 using two background variables, x_1 and x_2.

Insert Figure 2

The structural regression coefficients of the MIMIC model are interpreted just as ordinary partial regression coefficients. They are presented in a standardized form, except for dummy background variables where the coefficients will represent the expected standard deviation change in the factor when the dummy variable changes from one category to the other (e.g., from male to female). In these MIMIC analyses, the achievement variables will be treated as continuous, normally distributed variables despite their small numbers of scale steps and possible non-normality. Experience has shown that the estimates are rather robust to such deviations from normality. In order to decide on the number of factors that are important in the MIMIC modeling, initial factor analyses were performed on the achievement variables alone. Specific factors contributing less than 5% to the reliable variance were dropped before turning to MIMIC analysis. The MIMIC analyses were carried out in the LISCOMP computer program (Muthén, 1987).

Subgroup Means Estimated from the MIMIC model

The MIMIC model shows the influence of background variables on the factors as partial regression coefficients. It is also of interest to use the estimated model to compute estimated means for the achievement variables. In this way, mean differences in observed variables can be studied for subgroups corresponding to key NAEP reporting variables, such as gender and ethnicity, providing a more direct comparison between the two ways of describing the data.

The subgroup mean differences will be displayed graphically in line with Figure 1. Each graph corresponds to two subgroups to be compared, e.g., males and females. On the horizontal axis the estimated mean and variance for each of the two subgroups are used to plot an estimated distribution of general factor values, using normal approximations. The estimated means and variances are computed from the estimated model using the sample values for the background variables. The vertical axis refers to a specific content area and the graph displays the estimated regression lines of the content area score on the general factor, one line for each of the two subgroups. The two lines are determined by average parameter estimate values across the variables representing the content area. For simplicity, it is assumed that general and specific factors are uncorrelated. In this case, the two lines are parallel and their slope shows the influence of the general factor on the specific content area scores while the intercept difference shows a content area's estimated mean difference between the two subgroups, conditional on the general factor. This is the same as the estimated content-specific factor mean difference between the two subgroups. As discussed above, this difference is of primary interest because it shows the extent to which individuals in different subgroups differ in performance in a given content area despite having the same overall (general factor) score. The results will be presented in the scale of estimated standard deviations of the reliable portions of the observed variable variances. This standard deviation is obtained from the conditional variance given the background variables as estimated by the MIMIC model. Graphs will only be shown if "practically significant" deviations from unidimensionality are present, that is if the intercept difference is significant and exceeds 0.2 of this standard deviation, corresponding to a "small effect size" in ANOVA terms (a medium effect size is 0.5, and a large effect size is 0.8).

Results

The results of these analyses will be reported in three steps. First, the percentage variance contributed by the specific factors will be presented. Second, the structural regression coefficients will be given. Third, graphs for estimated subgroup means will be presented for content- and format-specific sets of items conditional on the general factor.

Results for the Measurement Part

The estimates for the measurement part of the structural (MIMIC) modeling will be described first. The percentages of specific factor variances are given in Table 5 below. It is seen that statistically significant deviations from unidimensionality are obtained with respect to three specific factors for grade 12 and four specific factors for grade 8. The percentages for these specific factors are in some cases sizable, ranging from 5-26% of the reliable portion of the observed variable (testlet) variation. For grade 12, the largest contributions are obtained for Data Analysis & Statistics in constructed-response format, Algebra in multiple-choice format, and Data Analysis & Statistics in multiple-choice format. For grade 8, the largest percentages of specific factor variance contributions are obtained for Geometry in constructed-response format, Geometry in multiple-choice format, and Measurement in multiple-choice format.

Insert Table 5

In order to compare these results with the content-factor analysis of 1990 NAEP math data by Rock (1991) and correlations among the NAEP scores for content areas, it is of interest to also present the correlations among the five content areas as deduced from the model (see appendix). These are given in Table 6. The correlations are somewhat higher than the values obtained in the Rock analysis for the 1990 test and are in line with the hypothetical examples shown at the end of the appendix. It is noteworthy that even with such high correlations differential subgroup differences can be found for the different factors as seen in the next section.

Insert Table 6

Results for the Structural Regressions (MIMIC Model)

Table 7 shows the grade 12 estimated coefficients for the set of regressions of the factors on the background variables. Many of the background variables show significant partial effects on several factors. The amount of variance (R^2) in each factor explained by the background variables is shown at the bottom of the table. The variation in the general factor is reasonably well explained by the background variables as indicated by the R^2 value of 49%.

Insert Table 7

It is interesting to compare the estimates in the general factor column with the 1992 NAEP report for overall proficiency. While the Table 7 MIMIC model refers to partial effects of a background variable given other background variables, the NAEP report refers to marginal effects for one background variable at a time. The marginal effect for a background variable is the result of interactions of this variable with other background variables and is not easily interpreted. Following are three Table 7 examples of differences in the outcomes of these two ways of reporting. For gender, the MIMIC model shows a significantly lower value for females given other background, while the NAEP report does not show a significant gender effect. It is not clear how the significant gender effect turns insignificant marginally. For Asian ethnicity, the reverse holds: the MIMIC model does not show a significant partial effect compared to Whites while the NAEP report shows a significant marginal effect. In this case, the interpretation may be that more Asians than Whites take advanced math courses, reducing the Asian effect when controlling for such course taking in the MIMIC model. In fact, while about the same percentage of Asians and Whites take second- or third-year Algebra (55%) and Geometry (57%), 16% of Asians take Calculus courses as compared to 5% of Whites and 28% of Asians take Trigonometry as compared to 19% of Whites. Finally, for school type, the MIMIC model shows a significant negative partial effect comparing Catholic schools to Public Schools, while the NAEP report shows a significant positive marginal effect. The estimates from the MIMIC model can also be used to describe marginal effects as described in the methods section. For example, the MIMIC-estimated marginal effect of Catholic schools versus Public Schools is clearly positive as in the NAEP report. This rough correspondence between the two approaches should hold for all background variables.

The specific-factor columns of Table 7 have a more complex interpretation because these factors refer to performance on content- and format-specific test items controlling for overall test performance (general factor value). A content- and format-specific factor may be seen as residual variation which describes a skill that goes beyond the general math test-taking skill. Such factors may correspond to content- and format-specific learning of new topics involving definitions, new concepts, and new procedures, and high values may correlate with high degrees of opportunity-to-learn for such specific topics. The specific factors M-Geom and C-Geom may be seen as validated by the strong specific-factor effects from Geometry and Trigonometry course taking as compared to not taking such courses and the specific factor M-Algebra may be seen as validated by the strong specific-factor effect from Calculus course taking. It is true that the students taking such advanced courses are on the whole more able at math, reflecting a selection phenomenon. The selection effect is, however, largely accounted for by the strong general factor effects seen for these course-taking categories and the specific-factor effects describe difference beyond such a general advantage.

The estimates in the M-Algebra specific-factor column for the Ethnicity background variables are noteworthy. They indicate that Blacks, Hispanics and Asians all have significantly higher M-Algebra values than the reference group of Whites (see also the Geometry columns for similar results). While Asians are significantly ahead on the specific M-Algebra factor, they are not significantly ahead of Whites on the general factor, other background variables held constant. This is an example of the multidimensional factor model being able to point to components of subgroup differences that are overlooked in terms of overall performance. The specific-factor finding is perhaps due to differences in opportunity-to-learn as a function of different course-taking choices. This Asian-White analysis result is relatively easy to describe. For Black and Hispanics, however, the M-Algebra advantage, i.e., the White disadvantage, is at first puzzling given their strong general-factor disadvantage relative to Whites. This can be understood by describing the situation as the White advantage on the general factor not leading to a fully comparable M-Algebra performance advantage, so that the model needs to moderate the White general-factor advantage by a lesser M-Algebra effect for Whites than for Blacks and Hispanics. This type of reasoning may also explain the two negative effects in the M-Data column for Alg-Calc course taking.

The possibility of differential effects of background on the different factors is an interesting feature of the multidimensional MIMIC model which makes for a richer representation of the data. Examples of differential and even opposite effects are found with respect to both content and format factors. For example, the partial effect of being female is significantly negative for the general factor, while significantly positive for the Algebra-specific factor in multiple-choice format and the Data Analysis-specific factor in constructed-response format. The partial effect of Asian versus White is small and insignificant for the general factor but large for the M-Geom and M-Algebra factors. In terms of format differences, Data Analysis & Statistics shows format differences for Females and for Blacks; in both cases performance in these groups is better on constructed-response items than multiple-choice items.

Table 8 shows the corresponding grade 8 MIMIC model estimates. In terms of differential effects of background on the factors, it is interesting to consider the background variable Gender. We find that with other background variables held constant, females are significantly higher than males on the general factor, but significantly lower on the Measurement-specific factor (in multiple-choice format). Geometry shows different relationships for the constructed-response format than for the multiple-choice format for females and for Blacks; here, females do better on the constructed-response format and Blacks do better on the multiple-choice format. It is also interesting to note that, as compared to grade 12, the Asian-White difference for Geometry has not yet developed. It should be noted, however, that the amount of variance explained in the specific factors is very low for grade 8.

Insert Table 8

Results for Subgroup Means Estimated from the MIMIC model

The following graphs show the estimates derived from the MIMIC model for subgroup mean differences in a given content area conditional on the general factor value. To limit space, only results for gender and ethnicity will be presented. As stated in the methods section, graphs are only presented if "practically significant" deviations from unidimensionality are present, requiring specific factor mean differences that are significant and at least 0.2 of a standard deviation of the reliable variation in the observed scores.

Gender comparisons

Grade 12 gender comparisons show no practically significant deviations from unidimensionality for any of the specific factors. Figure 3 shows a grade 8 gender comparison for the Measurement-specific factor in multiple-choice format. As shown in Table 5, this specific factor contributed approximately 20% of the reliable variation in the Measurement content area scores. The MIMIC results of Table 7 indicated that the partial effect of being female was positive, although rather small. The general factor distributions of Figure 3 also show that the marginal effect of being female is slightly positive. These results are in line with the 1992 NAEP report

(Mullis et al., 1992) for the overall math score viewing the overall math score in NAEP as a proxy for the general factor score. Conditional on the general factor score, however, males are ahead of females in Measurement performance. Had we not conditioned on the general factor, this gender difference in Measurement performance may not have been uncovered because the general factor dominates as a source of variation in the Measurement performance. The NAEP Data Almanac for 1992 math reflects this in that the gender mean difference is not significant and is only about 0.1 of a standard deviation. This female Measurement disadvantage may be seen as "unrealized potential" among females. While females do as well as males on the overall test, they fall behind in this particular area. It may be noted that the gender effect for Geometry is smaller than for Measurement (about 0.13 of a standard deviation as opposed to about 0.20).

Insert Figure 3

Figures 4 and 5 show the effects of different item formats. These figures compare male and female grade 12 performance on Data Analysis & Statistics, showing that in comparison to males, the constructed-response format suits females better than the multiple-choice format. While neither graph shows a large specific-factor difference, the reversal from a male advantage in Figure 4 (multiple choice) to a female advantage in Figure 5 still makes these two figures noteworthy.

Insert Figures 4 and 5

Ethnicity comparisons

Figures 6 and 7 show grade 12 Asian-White comparisons for Geometry (multiple-choice) and Algebra (multiple-choice). In both cases, Asians are ahead of Whites on the general factor and, conditional on the general factor, further ahead on Geometry and Algebra in multiple-choice form. The general factor difference in these two cases is rather small, less than 0.2 of a standard deviation. In contrast, the multidimensional MIMIC model is able to show that there are strong Asian-White differences with respect to specific Geometry and Algebra content and format, almost 0.4 and 0.6 of a standard deviation, respectively. As discussed in connection with Table 7, these differences may have to do with Asians taking more advanced courses than Whites. These differences may not show up as strongly in the observed scores because the specific factors only account for 12 and 16%, respectively of the reliable variances (see Table 5), the remainder corresponding to the dominant general factor variance. In this connection it is interesting to note what this finding says about the influence of test content on subgroup differences: had the 12th grade math test had more Geometry and Algebra content, the overall Asian-White difference would have been larger.

Insert Figures 6 and 7

Figure 8 shows a grade 12 Black-White comparison for Data Analysis & Statistics (multiple-choice) indicating a conditional advantage for Whites. It is noteworthy that despite such a strong White advantage for the general factor, this cannot fully explain the White advantage on these types of items. The specific-factor difference may have to do with lack of opportunity-to-learn for Blacks as compared to Whites for Data Analysis & Statistics type items.

Insert Figure 8

Figure 9 shows a grade 12 Black-White comparison for Algebra in multiple-choice format indicating a reversal in the comparisons of the two subgroups for the general versus the specific factors. The Black specific-factor advantage was mentioned in connection with the Table 7 results. The White general-factor advantage is not realized for these types of items. Perhaps this is due to there being only a small degree of overlap in the two general-factor distributions, so that the data supporting the two lines come mostly from high-performing Blacks and low-performing Whites.

Insert Figure 9

Figures 10, 11, 12 show grade 12 Black-Asian comparisons for Geometry (both formats) and Algebra (multiple-choice). In all cases, there is a specific-factor advantage for Asians which goes beyond the Asian general-factor advantage. Again, given that the subgroup differences pertain to more advanced topics, these advantages may have to do with opportunity-to learn differences.

Insert Figures 10-12

Figures 13 and 14 show grade 12 Hispanic-Black comparisons indicating a conditional Hispanic advantage for Data Analysis & Statistics (multiple-choice) and Geometry (constructed-response). The specific-factor difference is in both cases larger than the general-factor difference. One may note that the Data Analysis & Statistics finding is analogous to the White-Black comparison of Figure 8.

Insert Figures 13, 14

Figures 15, 16, 17 show grade 12 Hispanic-Asian comparisons. Figures 15 and 16 indicate a conditional Asian advantage for Geometry (multiple-choice) and Algebra (multiple-choice) as was the case in the White-Asian comparisons. Figure 17 shows an Asian disadvantage for Data Analysis & Statistics (multiple-choice) despite an Asian advantage for the general factor.

The interpretation of Figure 17 may be similar to that of Figure 5 in that the data supporting the two lines come mostly from high-performing Hispanics and low-performing Asians.

Insert Figures 15-17

Figures 18 and 19 show grade 8 Asian-White comparisons. Figure 18 shows that for the Measurement-specific factor in multiple-choice format there is a reversal in the effects for the general and the specific factors: Asians are ahead of Whites on the general factor, but Whites have conditionally higher values on Measurement. Figure 19 shows that for Data Analysis & Statistics in multiple-choice format an analogous reversal is seen. The NAEP Data Almanac shows that Asians obtain higher means in both content areas, but that the mean differences are insignificant.

Insert Figures 18, 19

Discussion

This paper has found multidimensionality in the 1992 NAEP math items. This has an impact on the description of subgroup differences. In several instances, the multidimensional description of subgroup differences was able to identify subgroup differences in content- and format-specific factors which were different from overall subgroup differences. This type of description indicates that the finding of highly correlated content-specific subscores does not necessarily suggest reporting only subgroup differences with respect to an overall score, but that reporting of conditional, content-specific scores may be used.

Studying subgroup differences with respect to specific factors may lead to a more "instructionally sensitive" way to analyze achievement data. Take, for example, the Asian-White difference with respect to Algebra shown in Figure 7. The specific-factor difference is almost 0.6 of a standard deviation (of the reliable part of the Algebra score) while the general factor difference is less than 0.2 of this standard deviation. The fact that Asian and White individuals with the same general factor value can differ this much with respect to what is specific to algebra raises the possibility of "unrealized potential" of the White student subgroup relative to the Asian subgroup. Another example is provided by the Figure 13 Hispanic-Black comparison for grade 12 Data Analysis & Statistics, suggesting that Blacks have unrealized potential relative to Hispanics. Such differences can reveal important educational process differences related to curricular emphases, differences in opportunity-to-learn, and the effects of differential course choices. It would be of interest to attempt to study such differences over time and to explain how they arise. As examples of other such specific factor differences worthy of further investigations one may also mention the Male-Female difference with respect to Measurement, the Asian-White difference with respect to Geometry, and the Black-White difference with respect to Data Analysis & Statistics. To understand these differences, however, it is likely that a much richer set of explanatory background variables is needed than was used here.

The differential subgroup differences for the different factor dimensions also clearly show how dependent subgroup differences are on the particular mix of content and format that is used for the test items. For example, in comparison to males, females appear to do relatively better on constructed-response items than multiple-choice items for Data Analysis & Statistics in grade 12 and Geometry in grade 8. This has implications for future developments of NAEP testing and the comparison of performance over time. One can expect a trend towards using more constructed-response items, reducing the reliance on the multiple-choice format. The particular content mix and the content weights may also change over time.
The 1992 math findings reported here replicate in some respects analyses of the 1990 NAEP math data (Muthén, 1991). In both cases, a MIMIC approach was taken, but analysis procedures were different in three regards. Due to the different BIB spiraling structures, the two data sets give rise to different ways of creating testlets. The 1990 data made it possible to analyze a set of testlets in seven replicate analyses of seven booklets, while in 1992 the analysis needed to be done simultaneously on all the 26 booklets. In the 1990 analyses no Asian-White or Black-Hispanic comparisons were made and no format-specific testlets or factors were formulated. Despite these differences, it is interesting to note that the 1992 grade 8 conditional Measurement disadvantage for females was also observed in analyses of the 1990 NAEP math data. Furthermore, the 1992 grade 12 Black-White comparison for Data Analysis & Statistics indicating a conditional advantage for Whites was also observed in analyses of the 1990 NAEP math data.

The latent variable technique used in this report provides a general methodology for data structures of the NAEP type. It gives flexibility for the researcher in that NAEP items and background variables are used without having to rely on the particular proficiency scores that are generated for NAEP reports. Conditioning variables are not used to generate scores. Such background variables can instead be incorporated in the analysis as done in the MIMIC model. This approach therefore provides a way to validate findings from regression analyses based on NAEP proficiency scores.

References

Carlson, J. E., & Jirele, T. (1992). Dimensionality of 1990 NAEP mathematics data. Paper presented at the annual meeting of the American Educational Research Association. San Franscisco, 1992.

Holzinger, K., J. & Swineford, F. (1939). A Study in Factor Analysis: The Stability of a Bi-factor Solution. Supplementary Educational Monographs, Number 48. The University of Chicago, Chicago, Illinois.

Mazzeo, J., Yamamoto, K., & Kulick, E. (1993).. Extended constructed-response items in the 1992 NAEP: Psychometrically speaking, were they worth the price? Paper presented at the annual meeting of the National Council on Measurement in Education. Atlanta. GA, April, 1993.

Mullis, I. V. S., Dossey, J. A., Owen, E. H., & Phillips, G. W. (1993). NAEP 1992. Mathematics Report Card for the Nation and the States. National Center for Education Statistics.

Muthén, B. (1987). LISCOMP. Analysis of linear structural equations with a comprehensive measurement model. Theoretical integration and user's guide. Mooresville, IN: Scientific Software.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Psychometrika, 54, 557-585.

Muthén, B. (1991). Issues in using NAEP mathematics items to study dimensionality, within-grade differences, and across-achievement grade growth. Presentation at the meeting of the Design and Analysis Committee of the National Assessment of Educational Progress, Watsonville, CA, 1991.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. Psychometrika, 42, 431-462.

Rock, D. (1991). Subscale dimensionality. In The NAEP 1990 Technical Report, Educational Testing Service.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. Psychometrika, 22, 53-61.

Appendix

When data are generated by a single dominant dimension and several minor dimensions, it is easy to settle for unidimensionality unless a special effort is made to find the additional dimensions. The following latent variable model is a useful tool for detecting such deviations from unidimensionality. The model is a classic "bi-factor" model (see e.g. Holzinger & Swineford, 1939) with one general factor and one specific factor for each observed variable. In the classic case, the specific factors are uncorrelated among themselves and with the general factor. This latent variable model will be referred to as a GS model (general-factor, specific-factor model). This model will be modified here to include covariates of the general and specific factors in which case all factors can be correlated as a function of their common dependence on the covariates. This modified GS model is the MIMIC model (multiple-indicators, multiple-causes model) used in the analyses of the paper. The modified GS model is a good vehicle for illustrating how multidimensional models may be mistaken for unidimensional models.

Consider the following GS model for ten observed variables y,

(1)

$$
\begin{aligned}
y1 &= G + \quad e1 \\
y2 &= G + \quad e2 \\
y3 &= G + \quad e3 \\
y4 &= G + \quad e4 \\
y5 &= G + \quad e5 \\
y6 &= G + \quad e6
\end{aligned}
$$

$$y_7 = G + S + e_7$$
$$y_8 = G + S + e_8$$
$$y_9 = G + S + e_9$$
$$y_{10} = G + S + e_{10},$$

where G and S are the general and specific factors, respectively, and e's represent measurement errors. For simplicity the above GS model has unit loadings everywhere. Consider next the structural regressions of the factors on a covariate x,

(2)

$$G = b_g x + r_g$$
$$S = b_s x + r_s$$

where the b's are regression coefficients and the r's are residuals. While the residuals are uncorrelated so that G and S are uncorrelated given x, the marginal correlation betweeen G and S is not zero. The point of involving a covariate x is the following. Using information on the y's alone, the correlation between G and S can only be identified under very restrictive specifications such as using fixed loadings. Adding information on x's, however, makes it possible to identify the structural regression coefficients and thereby allows G and S to correlate as a function of their common dependence on x. In such a model, the residual correlation for G and S is zero and no restrictive specifications are needed for the loadings. This appendix considers what happens in the conventional approach of analyzing only the y's and incorrectly applying a one-factor model when a modified GS model is the true model.

Assume for example that the first six y variables correspond to NAEP's Numbers & Operations items and the last four y variables correspond to Algebra items. Or, alternatively, that the first six y variables correspond to multiple-choice items for a certain content area and the last four y variables correspond to constructed-response items for the same content area. Using the first example, S corresponds to algebra-specific skills that go beyond the Numbers & Operations skills needed to solve the algebra items represented by y variables 7-10. A useful index of the degree to which the model deviates from unidimensionality is the specific factor variance ratio

(3) $\quad V(S) / \{ V(G) + V(S) + 2\, Cov\, (G, S) \}$,

where the covariance is zero in the classic GS model but possibly nonzero in the modified GS model with covariates. This ratio does not involve the variable-specific amount of measurement error variance. The proportion residual variance, or unreliability, in a y variable depends on the number of items used to form the testlets. It is advantageous that the ratio does not depend on this arbitrary choice. Here, reliability is defined as

(4)

$$\{ V(G) + V(S) + 2\, Cov\, (G, S)\} \;/\; \{ V(G) + V(S) + 2\, Cov\, (G, S) + V(e)\} ,$$

where for y variables 1-6 the terms V(S) and Cov (G, S) disappear.

The reliable part of the variation in the six Numbers & Operations variables is G and the reliable part of the four algebra variables is G + S. The correlation between these two reliable parts is

$$(5) \quad \{ V(G) + Cov(G, S) \} / \{ Sqrt[V(G)] \; Sqrt[V(G) + V(S) + 2 \, Cov(G, S)] \} .$$

In contrast to this correlation, the correlation between Numbers & Operations y variables and the Algebra y variables is attenuated because the measurement error variances add to the denominator of the expression above. The amount of attenuation depends on the reliability of the variables, which again depends on the number of items used to form the testlets.

The correlation given in (5) has further meaning. It is also the correlation that is obtained between the two factors of a two-factor, simple-structure confirmatory factor analysis model with correlated factors fitted to the y variables of the modified GS model. This is easily seen from (1) if factor 1 is defined as G and factor 2 is defined as G + S, letting variables 1-6 load on factor 1 and 7-10 load on factor 2. The fact that a correlated, two-factor model fits the GS model perfectly relates to hierarchical factor analysis transformations discussed in Schmid and Leiman (1957).

Using different choices of specific-factor variance ratio, G-S factor correlation, and variable reliability, a set of covariance matrices for the ten y variables were created and analyzed by a one-factor model. The values were chosen to be close to those seen in the NAEP analyses: the MIMIC-estimated grade 8 and 12 specific-factor variance ratios typically ranged from 0.1 to 0.3: grade 12 factor correlations for the general factor were 0.19 with the specific factor of Geometry (multiple-choice) and 0.14 with the specific factor of Algebra (multiple-choice): a typical value for the testlet reliability was around 0.4 while in Rock (1991) 0.7 was a more typical value given that more items per testlets were used (taking the square root of each the three reliability values given in Table A1 shows that they correspond to the one-factor standardized loadings of approximately 0.9. 0.8. and 0.7). The parameter values chosen for Table A1 give a 0.85-0.97 range for the two-factor correlation values (using equation 5) which is in line with the Rock (1991) findings for the five content areas of the 1990 NAEP math data as well as the corresponding results for the 1992 data given in this paper. Table A1 gives the chi-square values of fit for the misspecified one-factor model when analyzing a sample of n=500. The model has 35 df. In Table A1. the G. S factor correlation varies but for simplicity the specific-factor variance ratio given in Table A1 uses formula (3) with the G. S covariance set to zero.

Insert Table A.1.

It is seen that several combinations of parameter values give an acceptable fit to the incorrect one-factor model. implying that the power to reject this model is low. This occurs for low specific-factor variance ratio. low G-S factor correlation. and low variable reliability. One such case which appears to use typical parameter values based on the NAEP analyses. has specific-factor variance ratio of 0.2. G-S factor correlation of 0.2. and reliability of

0.5. The chi-square value is 24.71 in this case ($p=0.902$). The chi-square values are linear in the sample size so that with a sample of 1,000, a value twice as large would be obtained. Looking up the 5% critical value for 35 df.'s (approximately 49), one can also calculate that in this case a sample size of 992 would be required to reject the one-factor model at the 5% level. For this case, the correlation between the reliable parts of the two types of content variables is 0.91, i.e. a two-factor simple-structure confirmatory factor analysis model would have a factor correlation of 0.91 (this is independent of the reliability). Had a two-factor model been fitted to these data, such a high value is likely to also lead an investigator to maintain the one-factor model. The corresponding factor correlation for a specific-factor variance ratio of 0.1 is 0.96.

Table A1. Chi-square test values for misspecified one-factor model (35 df, n=500)

Reliability of y1 to y6 = 0.80

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.70	0.30	0.175	0.32		0.30

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.05	0.77	0.85	435.16	0.000
0.2	0.09	0.79	0.87	404.24	0.000
0.3	0.14	0.80	0.89	364.66	0.000
0.4	0.18	0.81	0.90	320.21	0.000
0.5	0.23	0.82	0.92	261.93	0.000

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.80	0.20	0.20	0.40		0.20

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.04	0.73	0.90	197.31	0.000
0.2	0.08	0.74	0.91	183.61	0.000
0.3	0.12	0.76	0.92	164.91	0.000
0.4	0.16	0.77	0.93	141.87	0.000
0.5	0.20	0.78	0.94	115.42	0.000

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.88	0.10	0.22	0.30		0.10

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.03	0.78	0.95	92.92	0.000
0.2	0.06	0.79	0.96	88.81	0.000
0.3	0.09	0.79	0.96	77.59	0.000
0.4	0.12	0.80	0.96	64.49	0.002
0.5	0.15	0.81	0.97	54.12	0.021

Reliability of y1 to y6 = 0.65

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.70	0.30	0.38	0.70		0.30

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.05	0.61	0.85	150.46	0.000
0.2	0.09	0.63	0.87	136.67	0.000
0.3	0.14	0.65	0.89	120.06	0.000
0.4	0.18	0.66	0.90	102.42	0.000
0.5	0.23	0.68	0.92	80.67	0.000

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.80	0.20	0.43	0.70		0.20

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.04	0.61	0.90	77.77	0.000
0.2	0.08	0.62	0.91	70.98	0.000
0.3	0.12	0.64	0.92	62.29	0.003
0.4	0.16	0.65	0.93	52.16	0.031
0.5	0.20	0.67	0.94	41.13	0.220

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.88	0.10	0.47	0.70		0.10

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.03	0.60	0.95	23.15	0.938
0.2	0.06	0.61	0.96	22.02	0.957
0.3	0.09	0.62	0.96	18.97	0.988
0.4	0.12	0.63	0.96	15.49	0.998
0.5	0.15	0.65	0.97	12.80	1.000

Reliability of y1 to y6 = 0.50

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.70	0.30	0.70	1.20		0.30

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.05	0.48	0.85	62.10	0.003
0.2	0.09	0.50	0.87	55.48	0.015
0.3	0.14	0.52	0.89	47.84	0.073
0.4	0.18	0.53	0.90	40.01	0.257
0.5	0.23	0.55	0.92	30.75	0.674

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.80	0.20	0.80	1.30		0.20

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.04	0.45	0.90	27.33	0.819
0.2	0.08	0.47	0.91	24.71	0.902
0.3	0.12	0.49	0.92	21.43	0.965
0.4	0.16	0.50	0.93	17.69	0.993
0.5	0.20	0.52	0.94	13.71	1.000

V(G)	V(S)	V(e1)	V(e7)		V(S)/[V(G)+V(S)]
0.88	0.10	0.88	1.20		0.10

r(G,S)	Cov(G,S)	Rel(y7-y10)	2-Fac Corr	Chi-sq	prob
0.1	0.03	0.46	0.95	8.30	1.000
0.2	0.06	0.48	0.96	7.85	1.000
0.3	0.09	0.49	0.96	6.70	1.000
0.4	0.12	0.50	0.96	5.41	1.000
0.5	0.15	0.52	0.97	4.43	1.000

Figure 1

Conditional representation of multidimensional scores

45

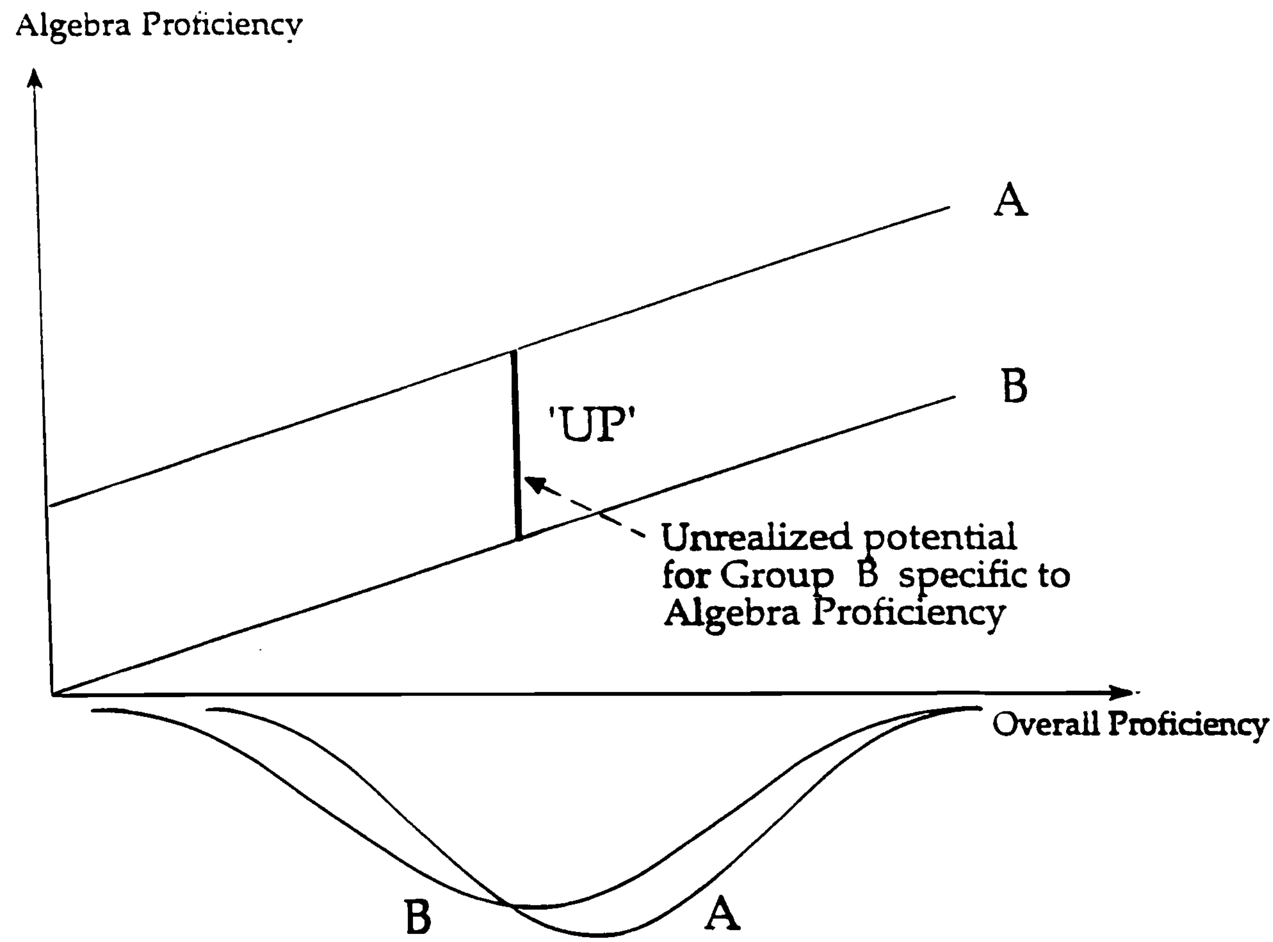

Algebra Proficiency
A
B
'UP'
Unrealized potential
for Group B specific to
Algebra Proficiency
Overall Proficiency
B
A

Figure 2

Path diagram for MIMIC model

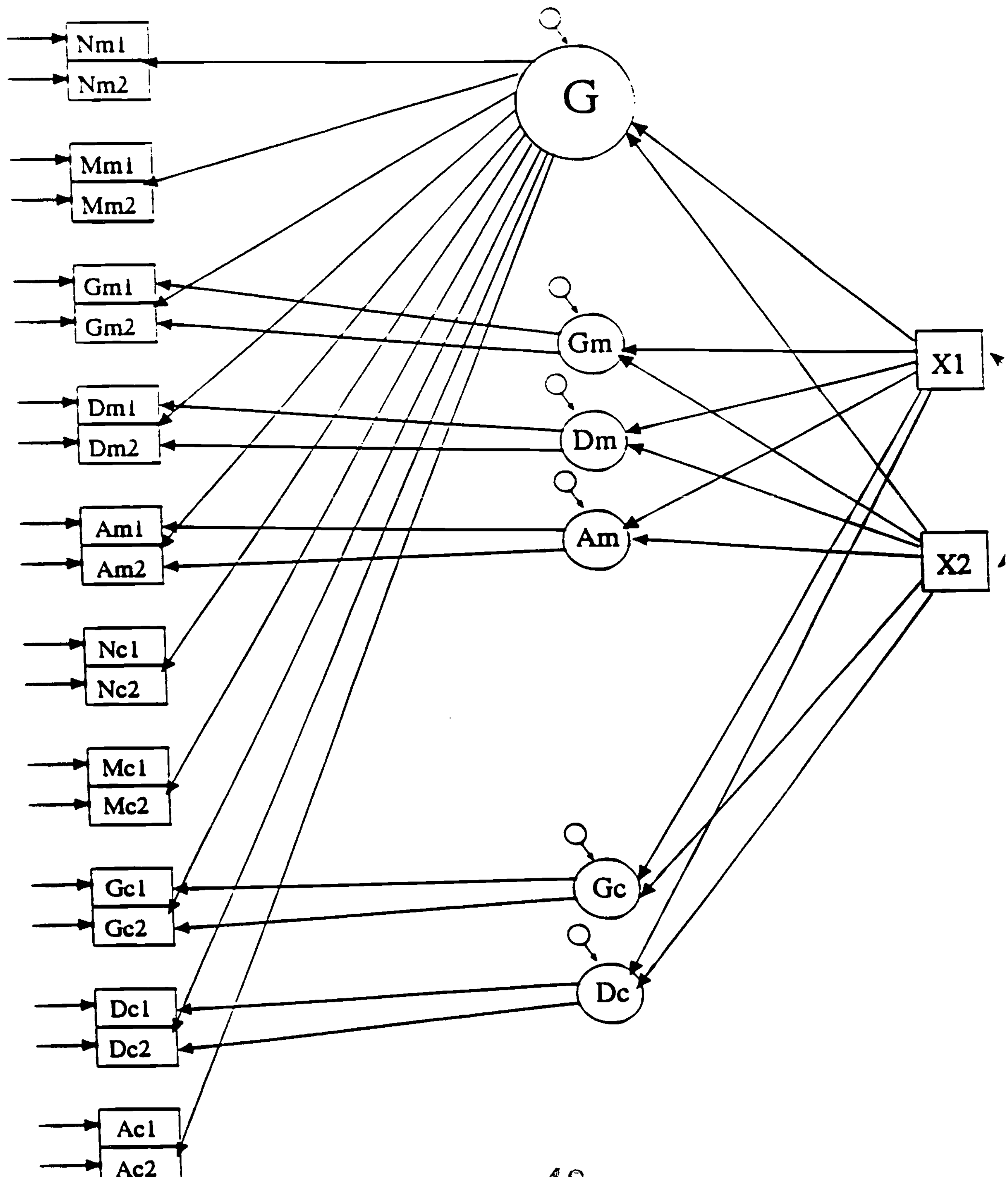
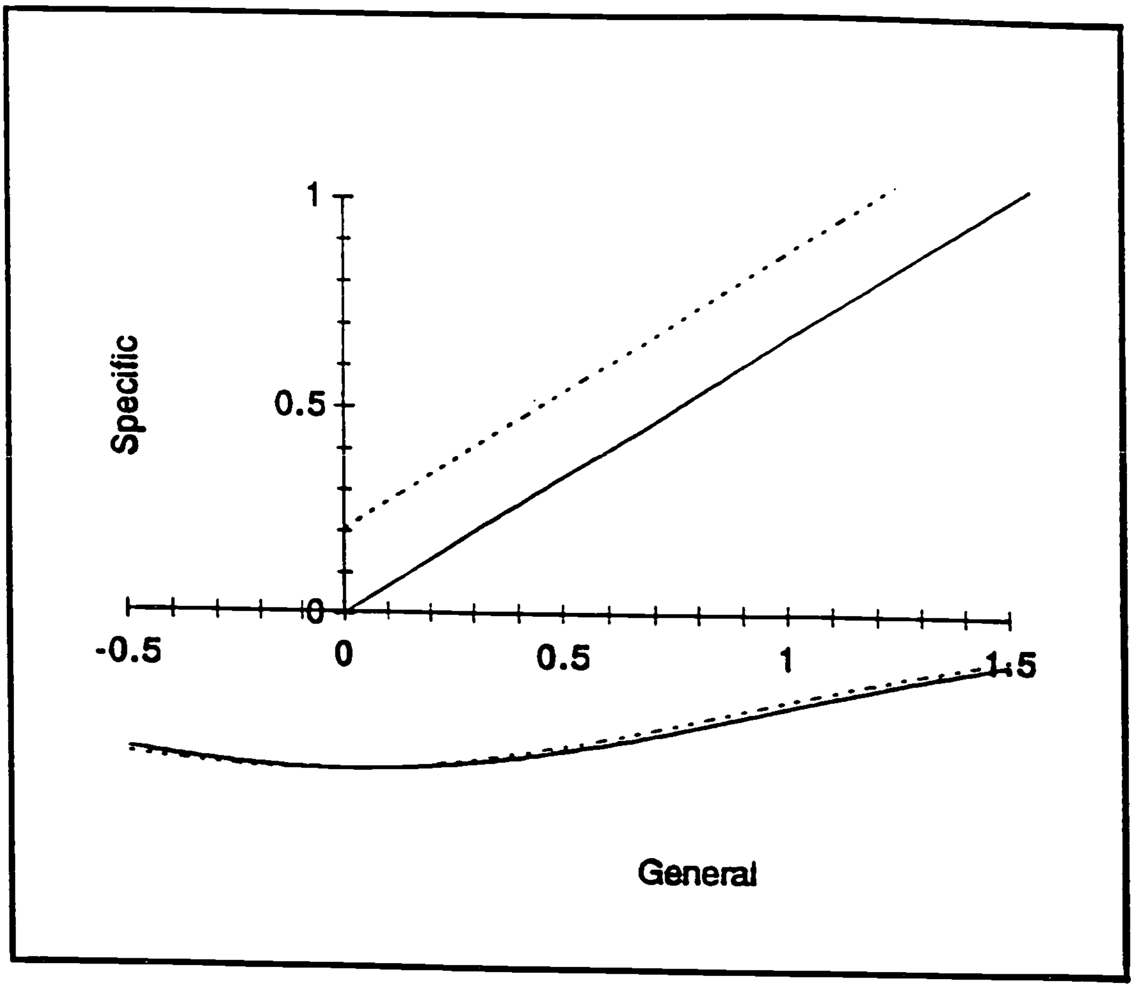

Nm1
Nm2
Mm1
Mm2
Gm1
Gm2
Dm1
Dm2
Am1
Am2
Nc1
Nc2
Mc1
Mc2
Gc1
Gc2
Dc1
Dc2
Ac1
Ac2
G
Gm
Dm
Am
Gc
Dc
X1
X2

Figure 3

Grade 8 gender comparison for the Measurement-specific factor in multiple-choice format

Female vs Male



Measurement

(Multiple-choice)

———— Female

- - - - - Male

Figure 4

Grade 12 gender comparison for the Data Analysis & Statistics-specific factor in multiple-choice format

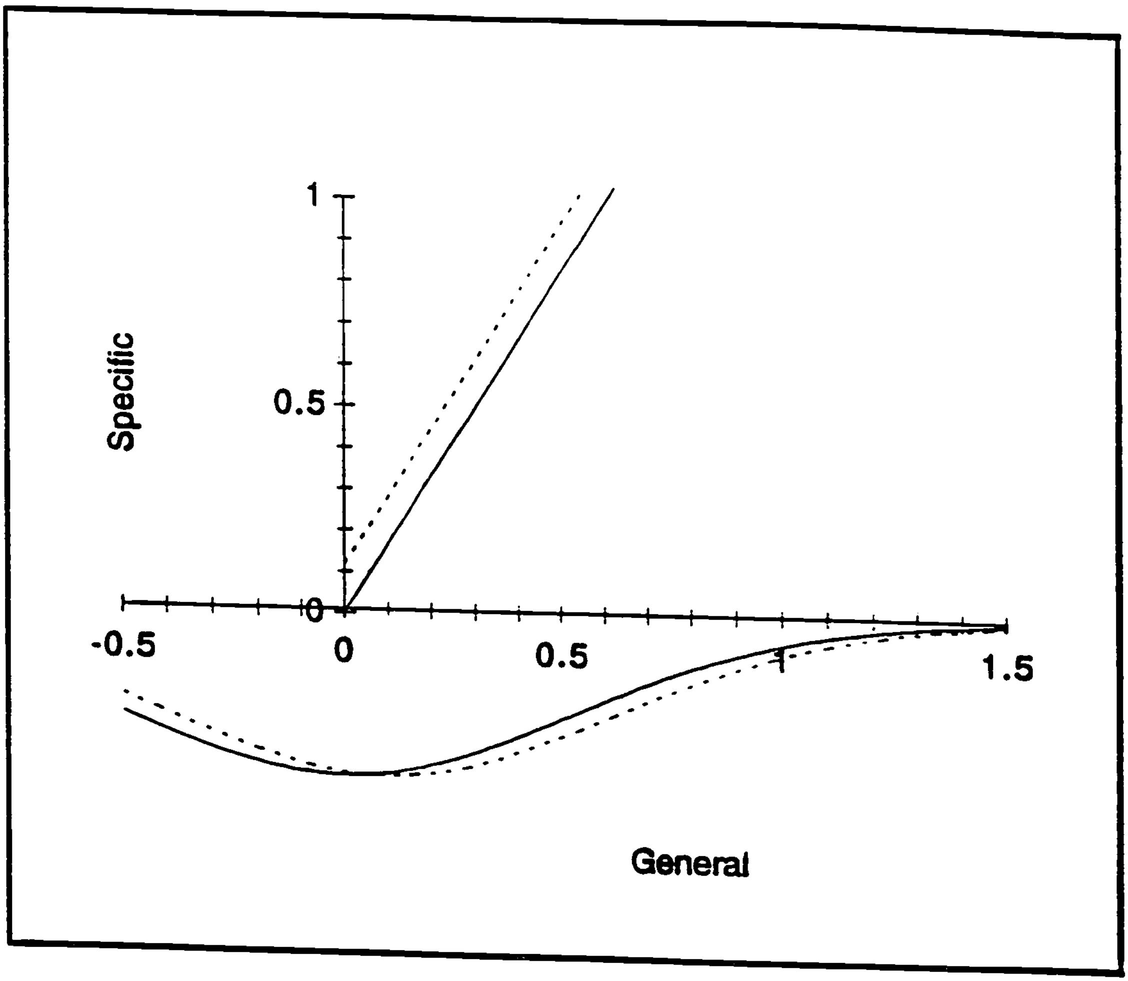
51

Figure 5

Grade 12 gender comparison for the Data Analysis & Statistics-specific factor in constructed-response format

NAEP '92 Grade 12

Female vs Male



Data analysis & Statistics

(Multiple-choice)

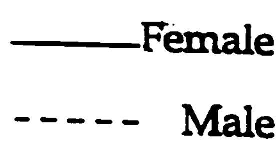
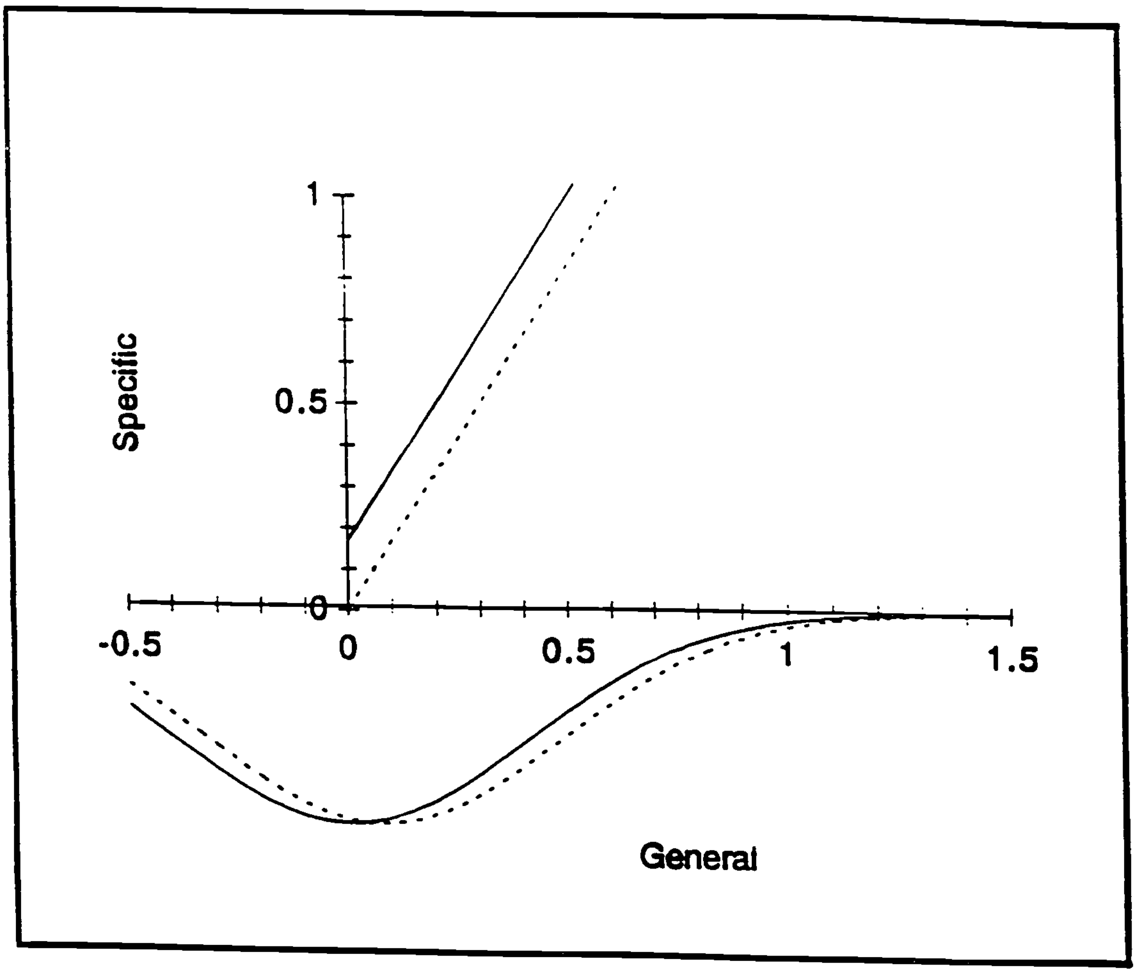


Figure 6

Grade 12 Asian-White comparisons for Geometry
in multiple-choice format.

NAEP '92 Grade 12

Female vs Male



Data analysis & Statistics

(Constructed-response)

——— Female

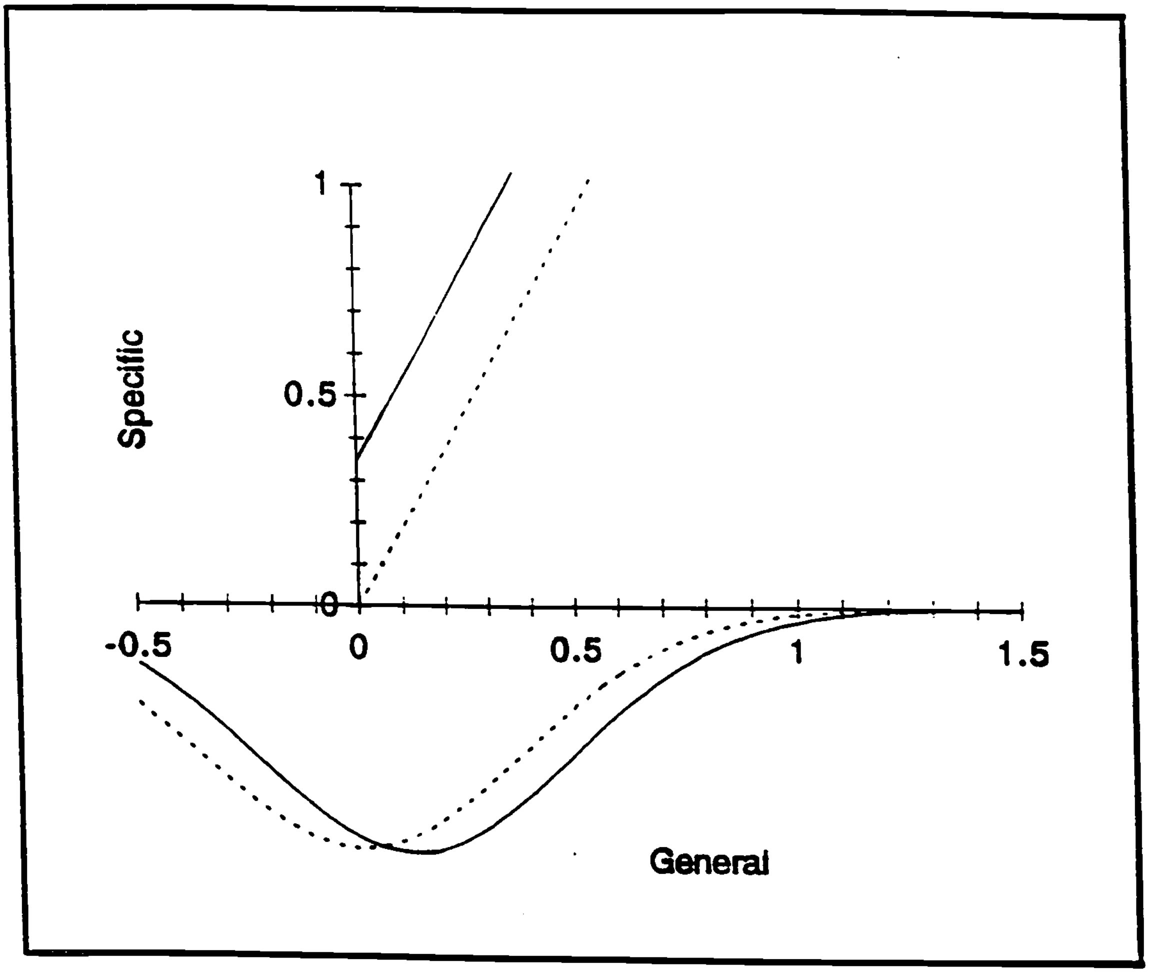
- - - - - Male

Figure 7

Grade 12 Asian-White comparisons for Algebra
in multiple-choice format.

NAEP '92 Grade 12

Asian vs White



Geometry

(Multiple-choice)

Asian

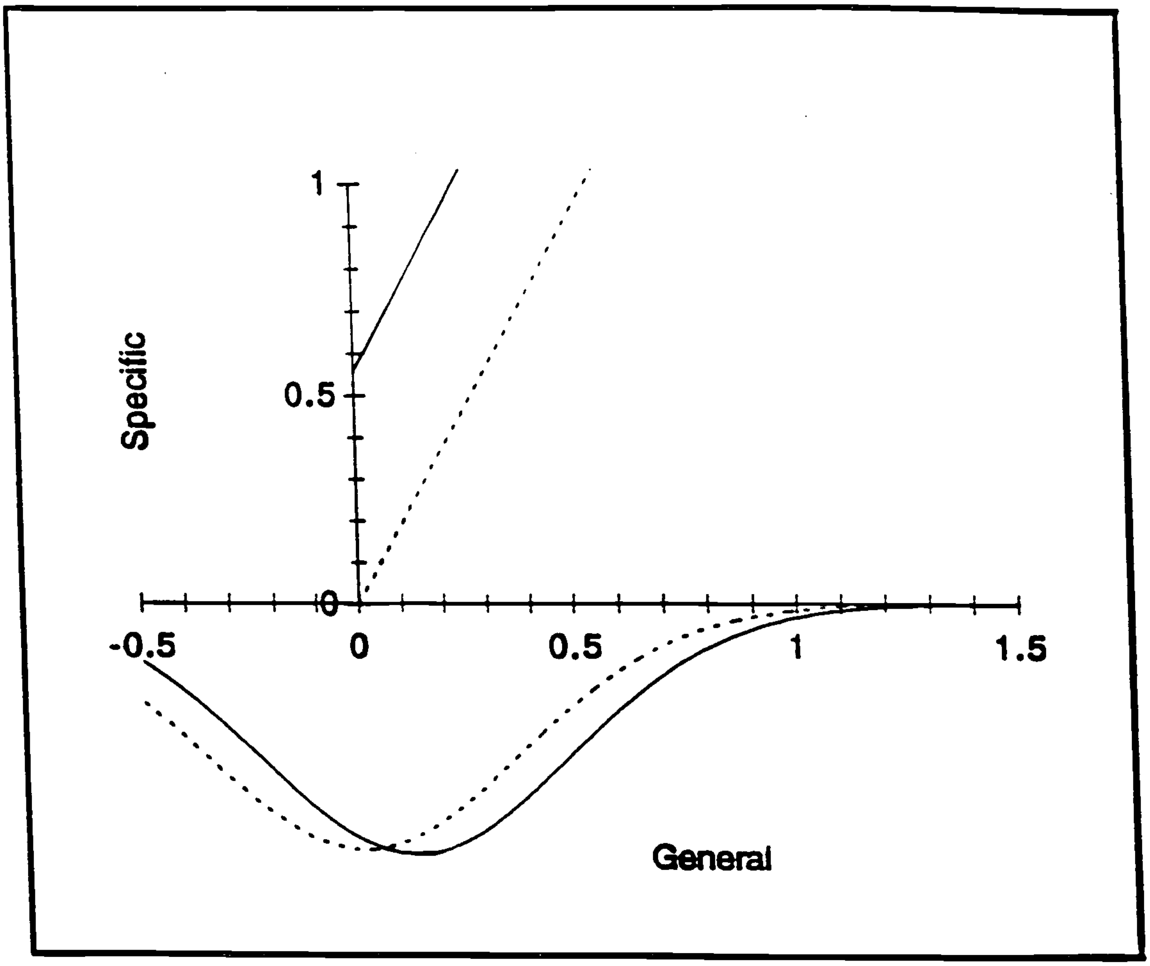
White

Figure 8

Grade 12 Black-White comparison for Data Analysis & Statistics in multiple-choice format.

NAEP '92 Grade 12

Asian vs White



Algebra

(Multiple-choice)

—— Asian

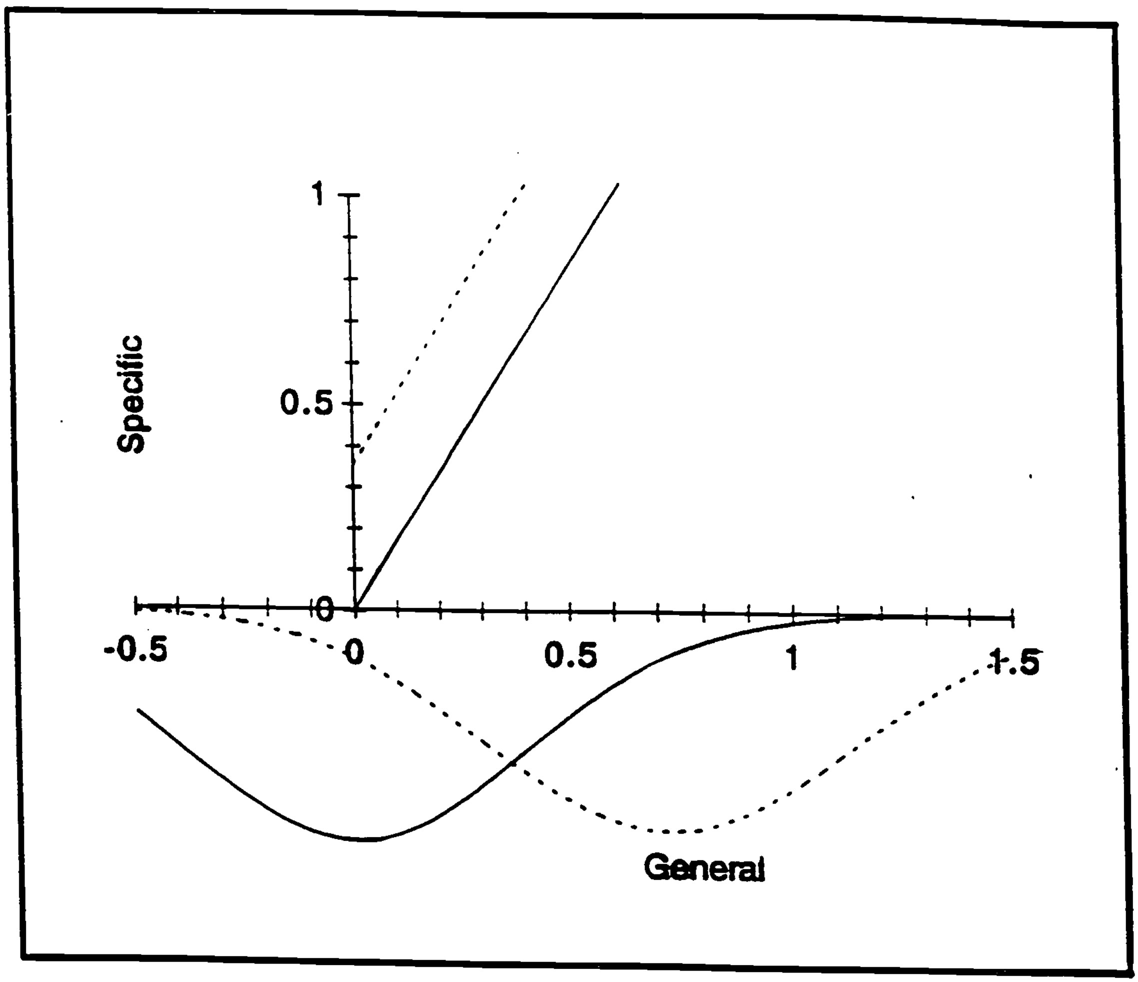
- - - - - White

Figure 9

Grade 12 Black-White comparison for Algebra in multiple-choice format

NAEP '92 Grade 12

Black vs White



Data analysis & Statistics

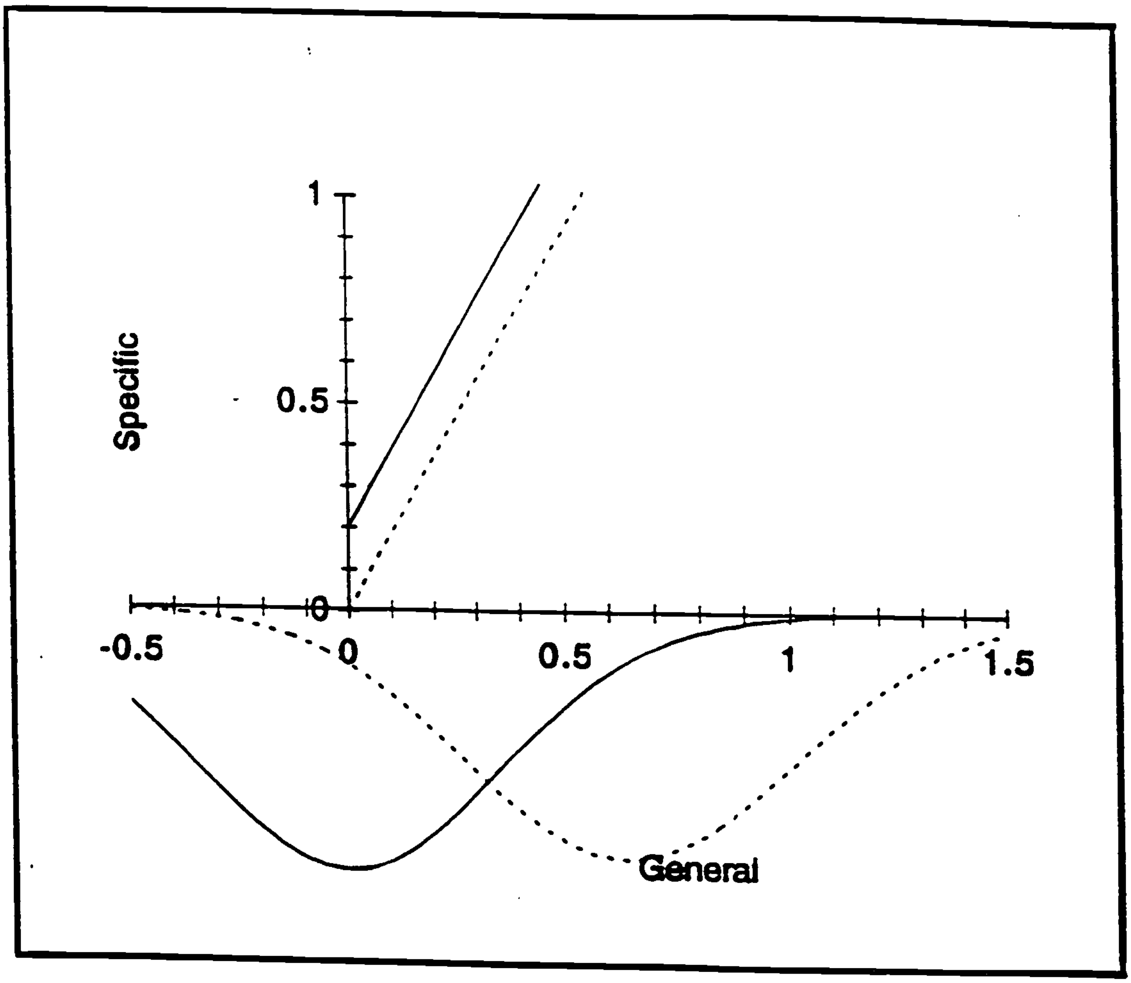
(Multiple-choice)

Figure 10

Grade 12 Black-Asian comparison for Geometry
in multiple-choice format

NAEP '92 Grade 12

Black vs White



Algebra

(Multiple-choice)

——— Black

- - - - - White

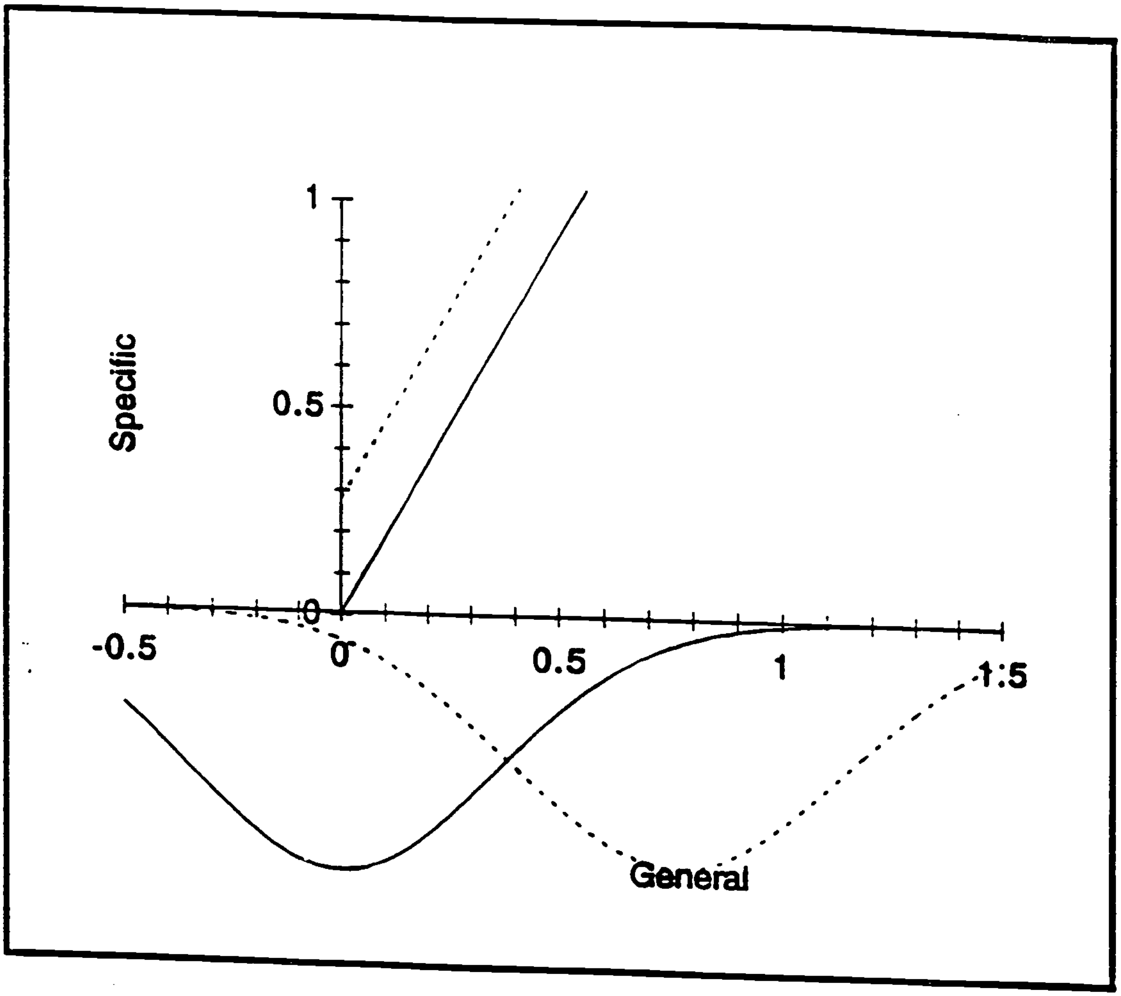
Figure 11

Grade 12 Black-Asian comparison for Geometry
in constructed-response format

64

NAEP '92 Grade 12

Black vs Asian



Geometry

(Multiple-choice)

Black (solid line)

Asian (dashed line)

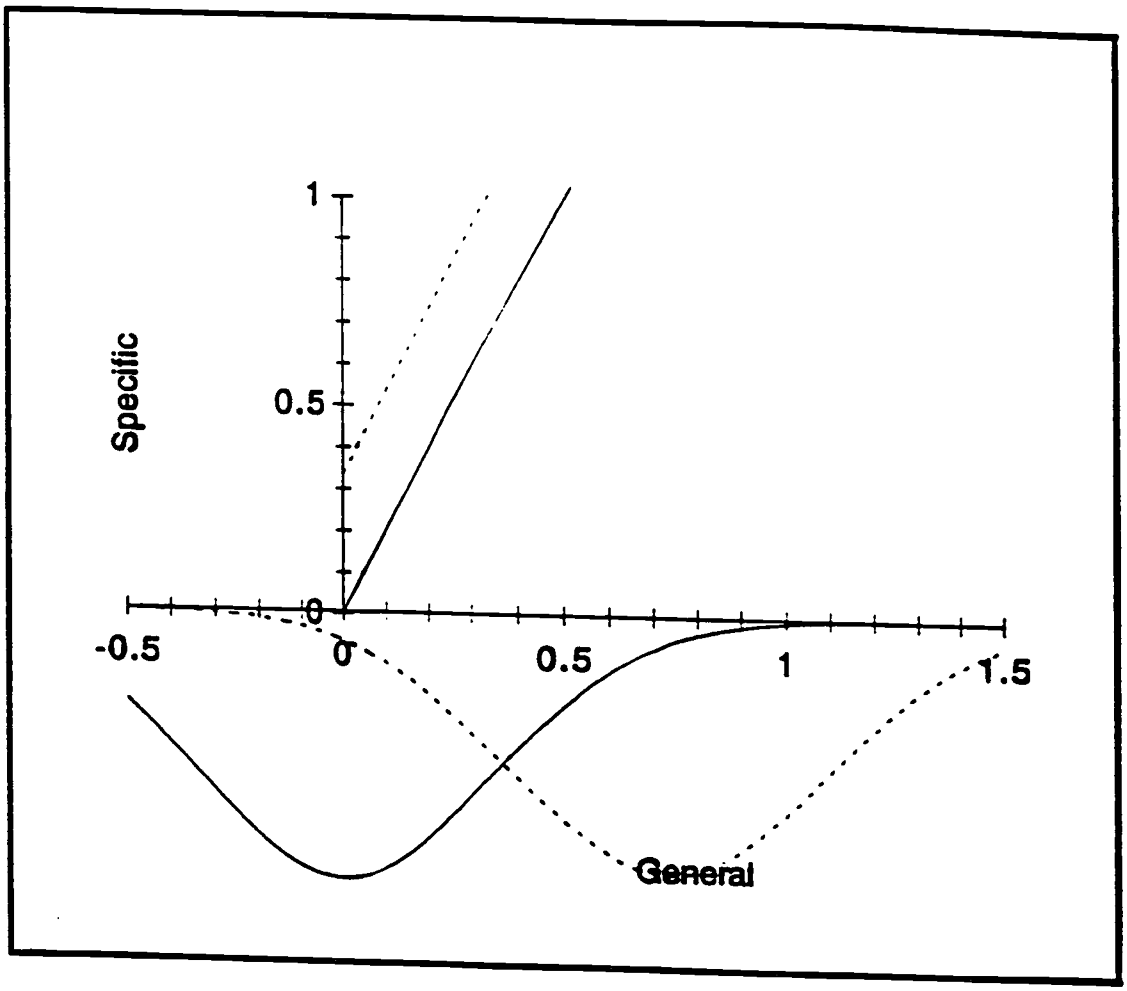
Figure 12

Grade 12 Black-Asian comparison for Algebra
in multiple-choice format

66

NAEP '92 Grade 12

Black vs Asian



Geometry

(Constructed-response)

Black

Asian

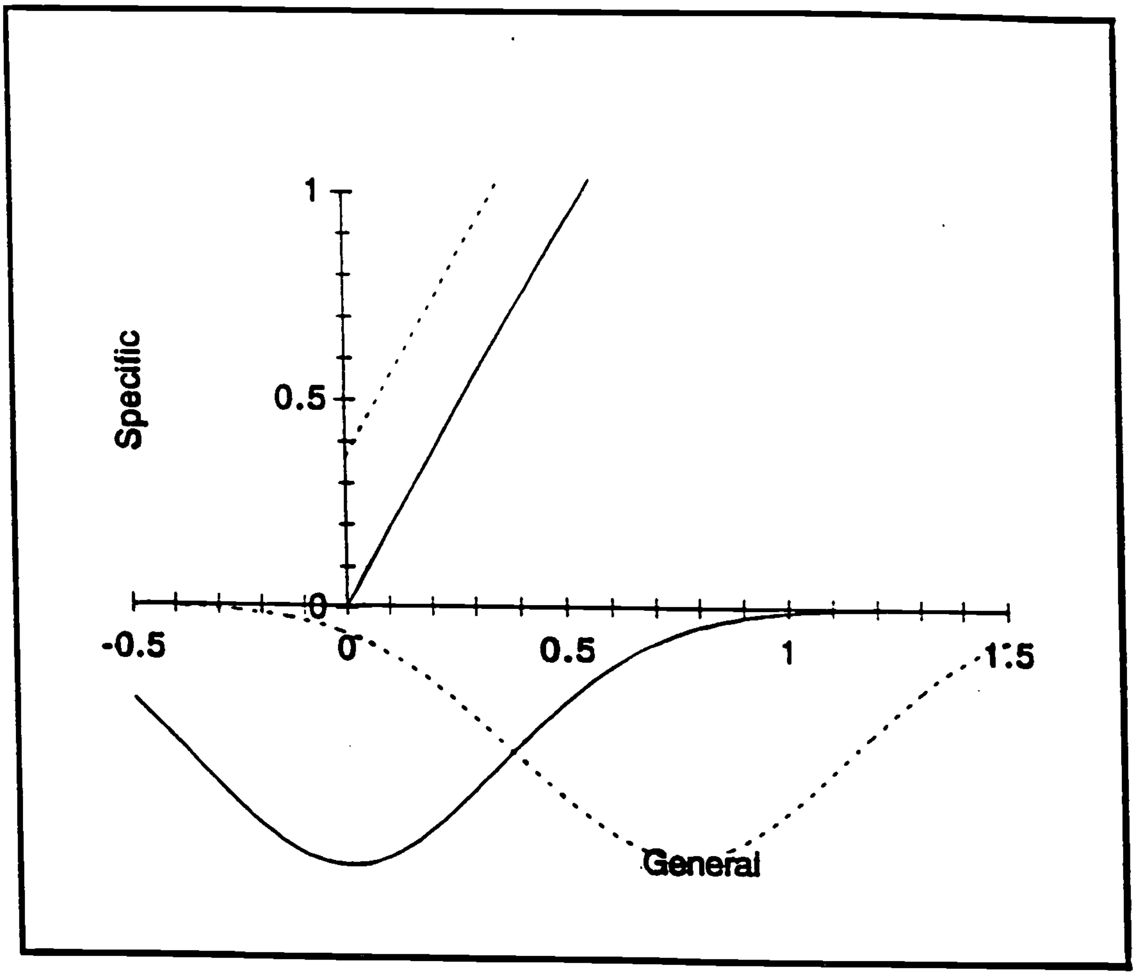
Figure 13

Grade 12 Hispanic-Black comparison for Data Analysis & Statistics in multiple-choice format.

68

NAEP '92 Grade 12

Black vs Asian



Algebra

(Multiple-choice)

Black

Asian

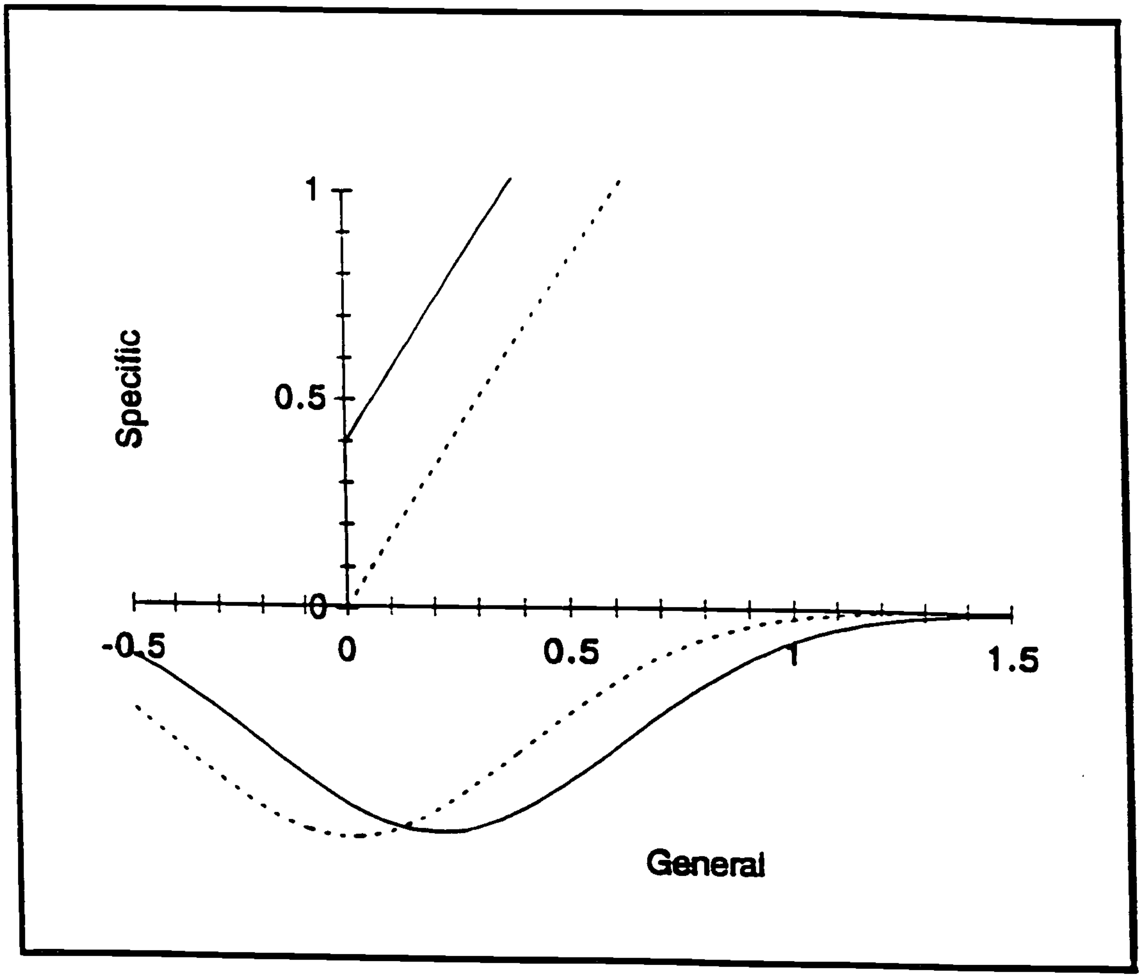
Figure 14

Grade 12 Hispanic-Black comparison for Geometry
in constructed-response format

70

NAEP '92 Grade 12

Hispanic vs Black



Data analysis & Statistics

(Multiple-choice)

——— Hispanic

- - - - - Black

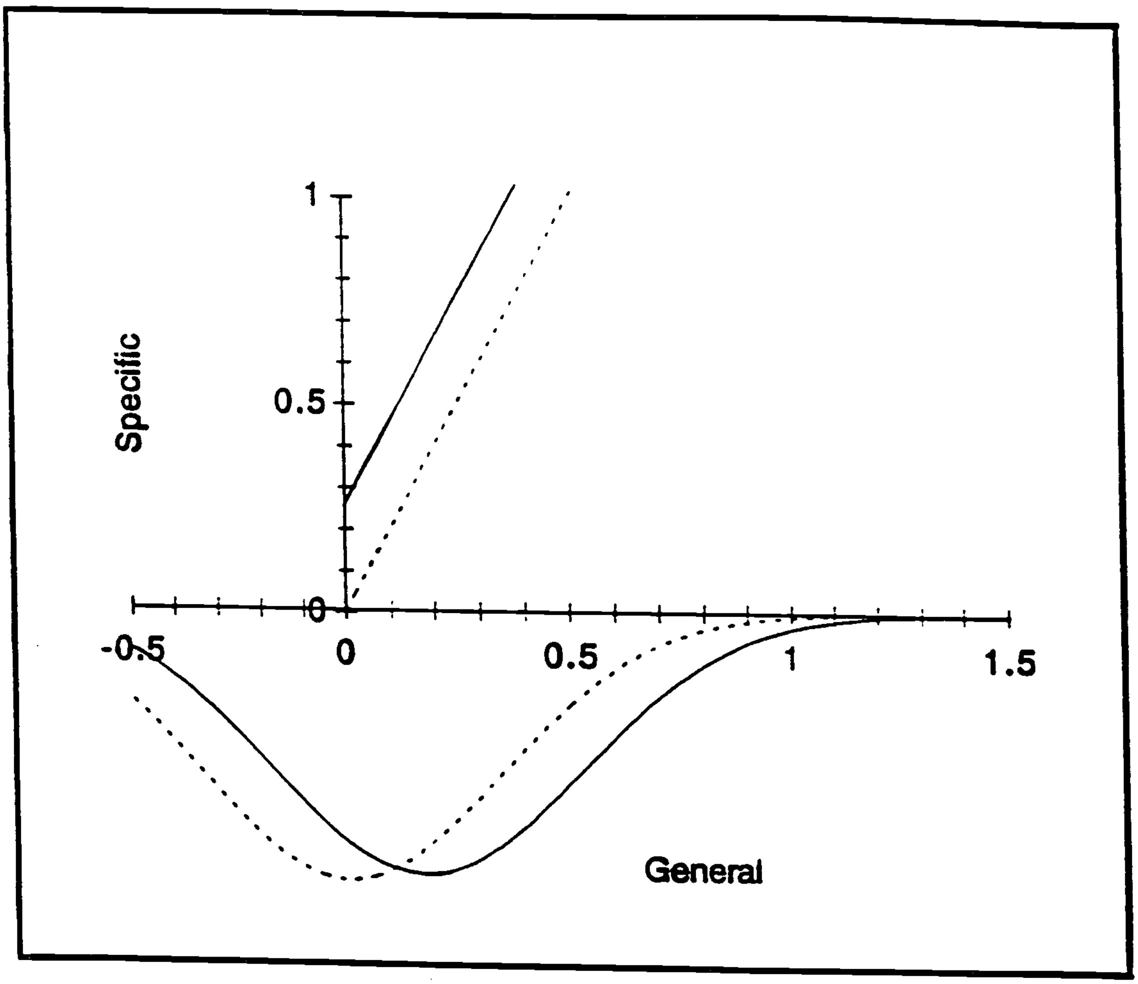
Figure 15

Grade 12 Hispanic-Asian comparison for Geometry
in multiple-choice format

72

NAEP '92 Grade 12

Hispanic vs Black



Geometry

(Constructed-response)

——— Hispanic

- - - - - Black

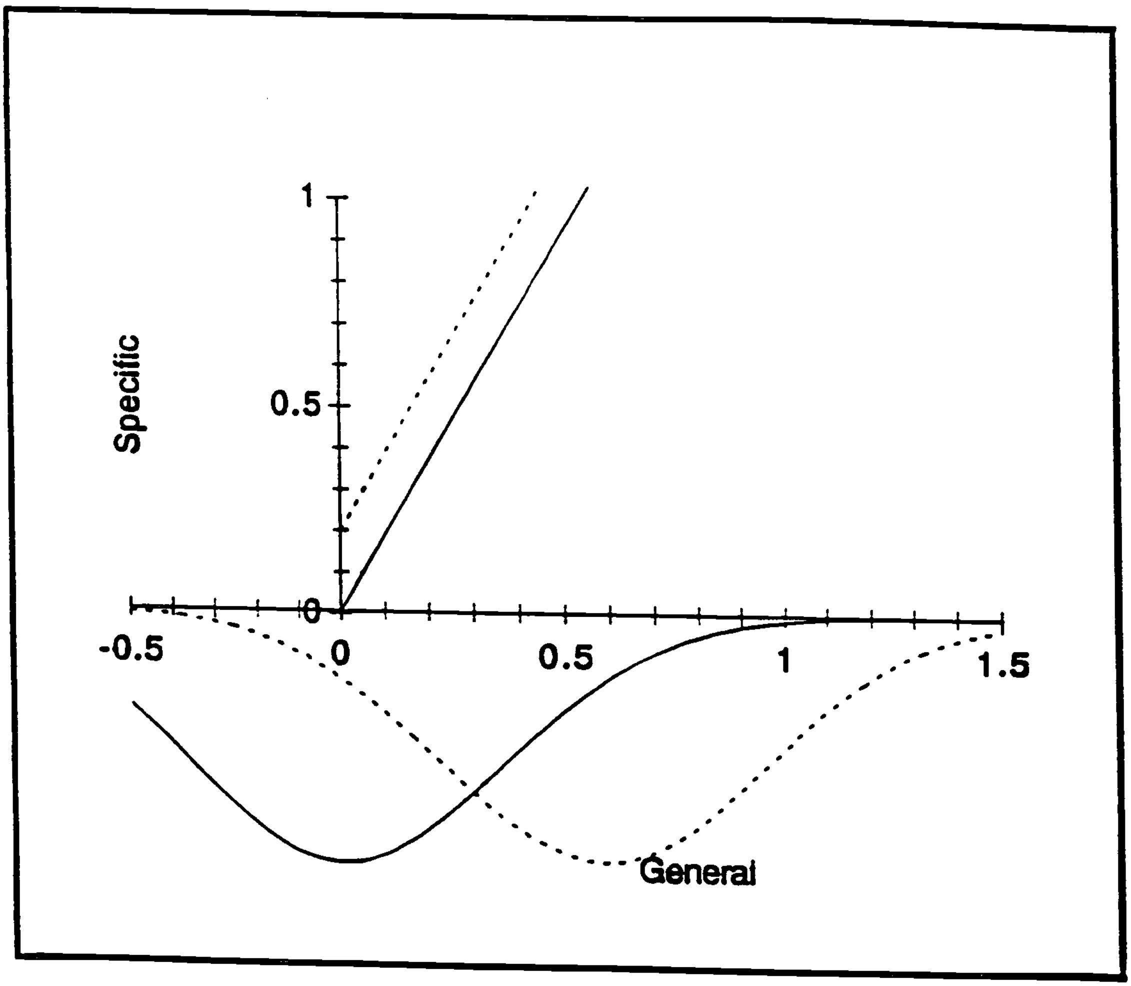
Figure 16

Grade 12 Hispanic-Asian comparison for Algebra
in multiple-choice format

74

NAEP '92 Grade 12

Hispanic vs Asian



Geometry

(Multiple-choice)

——— Hispanic

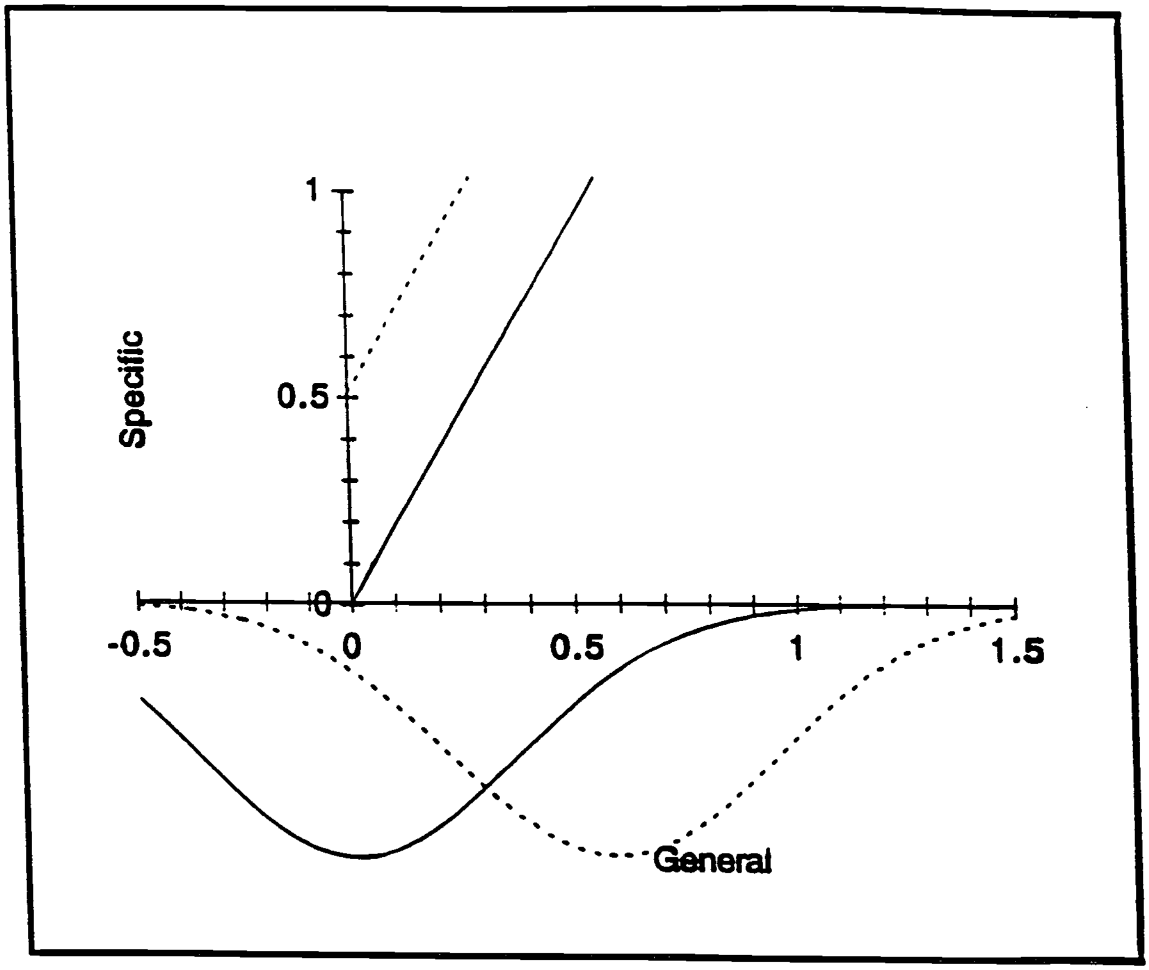
- - - - - Asian

Figure 17

Grade 12 Hispanic-Asian comparison for Data Analysis & Statistics in multiple-choice format

NAEP '92 Grade 12

Hispanic vs Asian



Algebra

(Multiple-choice)

——— Hispanic

- - - - - Asian

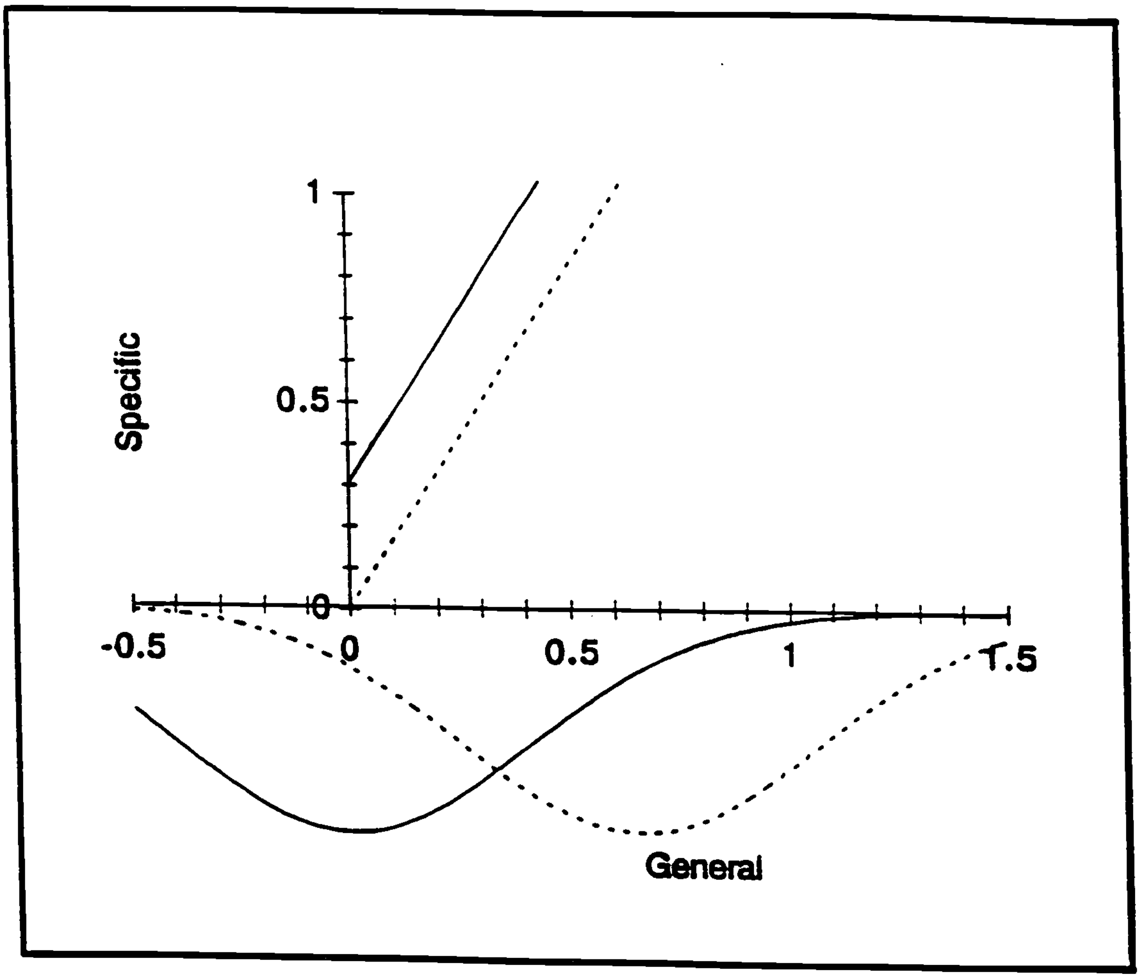
Figure 18

Grade 8 Asian-White comparison for the Measurement-specific factor in multiple-choice format

78

NAEP '92 Grade 12

Hispanic vs Asian



Data analysis & Statistics

(Multiple-choice)

——— Hispanic

- - - - - Asian

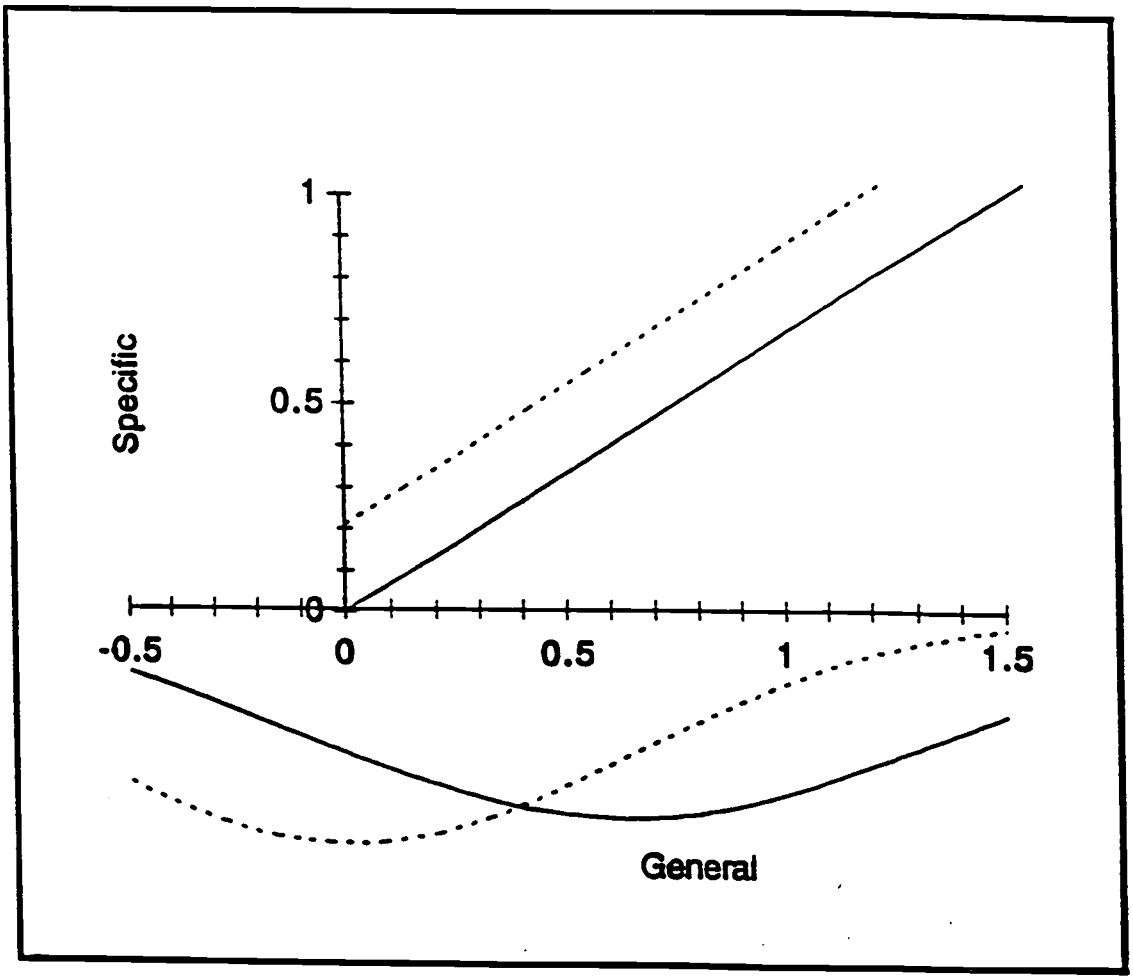
Figure 19

Grade 8 Asian-White comparison for Data Analysis & Statistics in multiple-choice format.

80

NAEP '92 Grade 8

Asian vs White



Measurement

(Multiple-choice)

—— Asian

----- White

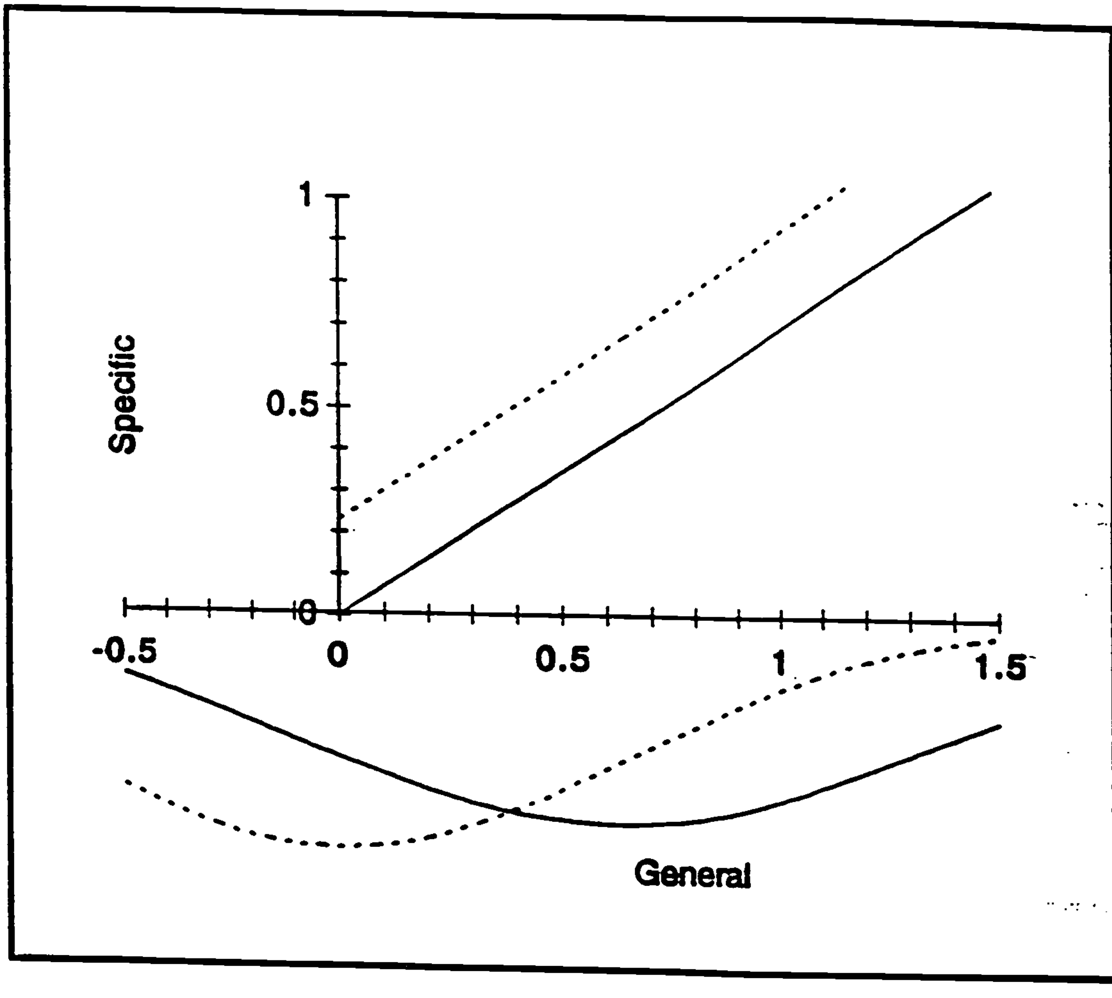
Table 1. Item content and format mix

NAEP'92 grade 12	Content	Num & Op	Measurement	Geometry	Data analysis	Algebra	Total
Format							
Multiple choice							
Number of items		29	18	20	17	32	116
	% of total	16.20%	10.06%	11.17%	9.50%	17.88%	100.00%
	% of content	25.00%	15.52%	17.24%	14.66%	27.59%	64.80%
	% of format	65.91%	64.29%	64.52%	58.62%	68.09%	
Short constructed response							
Number of items		15	10	10	11	11	57
	% of total	8.38%	5.59%	5.59%	6.15%	6.15%	100.00%
	% of content	26.32%	17.54%	17.54%	19.30%	19.30%	31.84%
	% of format	34.09%	35.71%	32.26%	37.93%	23.40%	
Extended constructed response							
Number of items		0	0	1	1	4	6
	% of total	0.00%	0.00%	0.56%	0.56%	2.23%	100.00%
	% of content	0.00%	0.00%	16.67%	16.67%	66.67%	3.35%
	% of format	0.00%	0.00%	3.23%	3.45%	8.51%	
Total							
Number of items		44	28	31	29	47	179
	% of content	24.58%	15.64%	17.32%	16.20%	26.26%	100.00%
	% of format	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

NAEP'92 grade 8	Content	Num & Op	Measurement	Geometry	Data analysis	Algebra	Total
Format							
Multiple choice							
Number of items		41	19	20	17	21	118
	% of total	22.40%	10.38%	10.93%	9.29%	11.48%	100.00%
	% of content	34.75%	16.10%	16.95%	14.41%	17.80%	64.48%
	% of format	70.69%	59.38%	55.56%	60.71%	72.41%	
Short constructed response							
Number of items		15	12	15	10	7	59
	% of total	8.20%	6.56%	8.20%	5.46%	3.83%	100.00%
	% of content	25.42%	20.34%	25.42%	16.95%	11.86%	32.24%
	% of format	25.86%	37.50%	41.67%	35.71%	24.14%	
Extended constructed response							
Number of items		2	1	1	1	1	6
	% of total	1.09%	0.55%	0.55%	0.55%	0.55%	100.00%
	% of content	33.33%	16.67%	16.67%	16.67%	16.67%	3.28%
	% of format	3.45%	3.13%	2.78%	3.57%	3.45%	
Total							
Number of items		58	32	36	28	29	183
	% of content	31.69%	17.49%	19.67%	15.30%	15.85%	100.00%
	% of format	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

NAEP '92 Grade 8

Asian vs White



Data analysis & Statistics

(Multiple-choice)

——— Asian

- - - - - White

Table 3. Layout of testlets in booklets arranged by response format within content areas
(entries are number of students)

				BOOKLETS																										
Testlets	FORMAT	BLOCK	CONTENT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Total
1	M	C	1	342									321			347	361						340					343		2054
2	M	D	1a	342	343									354				345						335					353	2072
3	M	D	1b	342	343									354				345						335					353	2072
4	M	E	1a		343	358									349		361		348						348					2107
5	M	E	1b		343	358									349		361		348						348					2107
6	M	G	1	342			353	335											348		341						353			2072
7	M	H	1a		343			335	346											353		325						343		2045
8	M	H	1b		343			335	346											353		325						343		2045
9	M	K	1					335			344	341						345					340		348					2053
10	M	L	1						346			341	321						348					335		354				2045
11	M	M	1							331			321	354						353					348		353			2060
12	M	O	1									341			349	347						325					353		353	2068
13	C	F	1			358	353									347		345		353						354				2110
14	C	I	1			358			346	331											341		340						353	2069
15	C	K	1					335			344	341						345					340		348					2053
16	C	M	1							331			321	354						353					348		353			2060
17	C	N	1								344			354	349						341					354		343		2085
18	M	D	2	342	343									354				345						335					353	2072
19	M	E	2		343	358									349		361		348						348					2107
20	M	H	2		343			335	346											353		325						343		2045
21	M	K	2					335			344	341						345					340		348					2053
22	M	N	2								344			354	349						341					354		343		2085
23	M	O	2									341			349	347						325					353		353	2068
24	C	E	2		343	358									349		361		348						348					2107
25	C	J	2				353			331	344						361					325		335						2049
26	M	D	3	342	343									354				345						335					353	2072
27	M	H	3		343			335	346											353		325						343		2045
28	M	K	3					335			344	341						345					340		348					2053
29	M	M	3							331			321	354						353					348		353			2060
30	M	O	3									341			349	347						325					353		353	2068
31	C	E	3		343	358									349		361		348						348					2107
32	C	F	3a			358	353									347		345		353						354				2110
33	C	F	3b			358	353									347		345		353						354				2110
34	C	F	3c			358	353									347		345		353						354				2110
35	C	J	3a				353			331	344						361					325		335						2049
36	C	J	3b				353			331	344						361					325		335						2049
37	M	C	4	342									321			347	361						340					343		2054
38	M	D	4	342	343									354				345						335					353	2072
39	M	E	4		343	358									349		361		348						348					2107
40	M	G	4	342			353	335											348		341						353			2072
41	M	I	4			358			346	331											341		340						353	2069
42	M	K	4					335			344	341						345					340		348					205
43	C	F	4			358	353									347		345		353						354				211
44	C	H	4		343			335	346											353		325						343		204
45	C	K	4					335			344	341						345					340		348					205
46	M	C	5	342									321			347	361						340					343		20
47	M	D	5	342	343									354				345						335					353	20
48	M	E	5		343	358									349		361		348						348					21
49	M	H	5		343			335	346											353		325						343		20
50	M	K	5					335			344	341						345					340		348					20
51	M	O	5									341			349	347						325					353		353	20

BEST COPY AVAILABLE

Table 2. Layout of testlets in booklets arranged by response format within content areas
(entries are number of students)

NAEP '92 grade

Testlets	FORMAT	BLOCK	CONTENT	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	Tot
1	M	C	1	341									333			318	314						340					345		19
2	M	D	1a	341	338									332				328						336					333	20
3	M	D	1b	341	338									332				328						336					333	20
4	M	E	1		338	331									328		314		331						349					19
5	M	H	1		338			340	331											326		340						345		20
6	M	K	1					340			326	356						328					340		349					20
7	M	L	1						331			356	333						331					336		331				20
8	C	C	1	341									333			318	314						340					345		19
9	C	E	1		338	331									328		314		331						349					19
10	C	F	1			331	354									318		328		326						331				19
11	C	G	1	341			354	340											331		344						341			20
12	C	N	1								326			332	328						344					331		345		20
13	C	O	1									356			328	318						340					341		333	20
14	M	D	2	341	338									332				328						336					333	20
15	M	E	2		338	331									328		314		331						349					19
16	M	G	2	341			354	340											331		344						341			20
17	M	H	2		338			340	331											326		340						345		20
18	M	N	2								326			332	328						344					331		345		20
19	M	C	3	341									333			318	314						340					345		19
20	M	D	3	341	338									332				328						336					333	20
21	M	E	3		338	331									328		314		331						349					19
22	M	G	3	341			354	340											331		344						341			20
23	M	H	3		338			340	331											326		340						345		20
24	M	K	3					340			326	356						328					340		349					20
25	M	O	3									356			328	318						340					341		333	20
26	C	F	3			331	354									318		328		326						331				19
27	C	J	3				354			319	326						314					340		336						19
28	C	M	3							319			333	332						326					349		341			20
29	M	D	4	341	338									332				328						336					333	20
30	M	E	4		338	331									328		314		331						349					19
31	M	H	4		338			340	331											326		340						345		20
32	M	I	4			331			331	319											344		340						333	19
33	M	M	4							319			333	332						326					349		341			20
34	C	F	4a			331	354									318		328		326						331				[illegible]
35	C	F	4b			331	354									318		328		326						331				19
36	C	H	4		338			340	331											326		340						345		20
37	M	C	5	341									333			318	314						340					345		19
38	M	D	5a	341	338									332				328						336					333	20
39	M	D	5b	341	338									332				328						336					333	20
40	M	E	5		338	331									328		314		331						349					[illegible]
41	M	H	5		338			340	331											326		340						345		20
42	M	I	5			331			331	319											344		340						333	19
43	M	J	5				354			319	326						314					340		336						19
44	M	K	5					340			326	356						328					340		349					20
45	M	O	5									356			328	318						340					341		333	20
46	C	F	5a			331	354									318		328		326						331				[illegible]
47	C	F	5b			331	354									318		328		326						331				19
48	C	G	5	341			354	340											331		344						341			20
49	C	L	5						331			356	333																	

Table 5. Average percentage contribution of specific factors to reliable testlet variation

Factor	Variance	T-value	% Contribution
NAEP '92 grade 12			
1. General	0.09	11.00	80.40
2. M-Measurement	0.00	–	–
3. M-Geometry	0.05	2.45	10.97
4. M-Data Analysis & Statistics	0.04	1.11	17.40
5. M-Algebra	0.06	4.07	13.27
6. C-Numbers & Operations	0.00	–	–
7. C-Geometry	0.03	0.48	5.53
8. C-Data Analysis & Statistics	0.10	2.74	19.30
9. C-Algebra	0.00	–	–
NAEP '92 grade 8			
1. General	0.84	21.22	79.05
2. M-Measurement	0.10	4.53	14.78
3. M-Geometry	0.10	3.56	23.47
4. M-Data Analysis & Statistics	0.06	2.13	11.44
5. M-Algebra	0.02	0.69	–
6. C-Numbers & Operations	0.04	1.28	7.35
7. C-Measurement	0.03	0.42	–
8. C-Geometry	0.25	8.49	25.53
9. C-Data Analysis & Statistics	0.00	–	–

M=Multiple choice
C=Constructed response

Table 4. Background variables used in the structural model (NAEP '92)

Sample Size		8963	8705
		% in Grade 8	% in Grade 12
1. Gender			
	*1 Male	51	49
	2 Female	49	51
2. Ethnicity			
	*1 White	67	69
	2 Black	16	17
	3 Hispanic	14	10
	4 Asian	3	4
3. Parents' Education (Student Reported)			
	1 Didn't Finish High School	9	8
	2 Grad From High School	25	22
	3 Some Ed After High School	20	26
	4 Grad From College	47	44
4. Type Of Community			
	1 Extreme Rural	8	11
	2 Disadvantaged Urban	10	13
	3 Advantaged Urban	11	12
	*4 Other (Non-Extreme)	71	64
5. School Type			
	*1 Public School	79	80
	2 Private School	8	7
	3 Catholic School	13	13
6. Algebra (Course Taking)			
	1 Pre-Algebra/Algebra	44	
	*2 No Algebra/Other	56	
7. Alg-Calc (Course Taking)			
	*1 Pre-Algebra/1st-Year Algebra/Not Studied		44
	2. 2nd/3rd-Year Algebra		52
	3 Calculus		4
8. Geom-Trig (Course Taking)			
	*1 Not Studied		26
	2 Geometry		56
	3 Trigonometry		18
9. School Program			
	*1 General		22
	2 Academic/College Prep		26
	3 Vocational/Technical		48
	4 Other/Omitted		4

Categories in the background variables are all dummy coded except for Parents' Education. For dummy-coded variables, effects are interpreted as the category in question compared to base category (marked *) of the variable.

Table 7. Standardized coefficients (& t-values) from the structural model (NAEP '92 grade 12)

	General	M-Geom	M-Data	M-Algebra	C-Geom	C-Data
Female	-0.140 (-6.53)	-0.008 (-0.14)	-0.214 (-1.78)	0.198 (2.62)	-0.026 (-0.19)	0.388 (3.30)
Ethnicity						
Black	-0.705 (-16.12)	0.275 (1.90)	-0.977 (-5.20)	0.608 (5.23)	-0.275 (-0.95)	-0.288 (-1.69)
Hispanic	-0.402 (-10.06)	0.489 (3.00)	0.050 (0.20)	0.302 (2.25)	0.711 (2.03)	-0.362 (-1.85)
Asian	0.015 (0.28)	0.673 (2.89)	-0.425 (-1.37)	1.099 (5.79)	0.734 (1.47)	-0.537 (-1.85)
Parents Ed.	0.107 (8.83)	-0.006 (-0.12)	0.050 (0.74)	0.025 (0.61)	0.087 (0.80)	-0.040 (-0.64)
TOC						
Rural	0.191 (5.53)	0.076 (0.55)	0.021 (0.13)	0.197 (1.66)	-0.048 (-0.12)	0.270 (1.48)
Disadv-Urban	-0.149 (-4.54)	0.072 (0.48)	0.547 (2.80)	0.105 (0.87)	-0.054 (-0.21)	0.099 (0.50)
Adv-Urban	-0.054 (-1.63)	-0.226 (-1.51)	0.318 (1.59)	-0.054 (-0.45)	0.175 (0.56)	0.181 (0.96)
School-Type						
Catholic	-0.135 (-4.12)	0.088 (0.61)	-0.144 (-0.74)	-0.088 (-0.76)	-0.426 (-1.40)	0.085 (0.45)
Private	0.097 (2.39)	-0.004 (-0.04)	-0.467 (-1.91)	0.471 (3.15)	0.284 (0.71)	-0.234 (-1.02)
Alg-Calc						
Algebra	0.394 (12.85)	-0.126 (-1.04)	-0.594 (-3.82)	0.136 (1.53)	-0.604 (-2.42)	-0.070 (-0.40)
Calculus	0.849 (12.66)	0.259 (1.08)	-0.869 (-2.54)	0.932 (4.95)	0.342 (0.68)	-0.459 (-1.38)
Geom-Trig						
Geometry	0.463 (13.25)	1.149 (9.04)	-0.010 (-0.03)	0.062 (0.64)	0.938 (3.48)	0.089 (0.62)
Trigonometry	0.595 (12.93)	1.218 (7.21)	-0.237 (-1.01)	0.420 (3.08)	0.918 (2.57)	-0.068 (-0.26)
School-Program						
Academic	0.422 (12.66)	0.024 (0.20)	-0.202 (-1.19)	0.264 (2.62)	-0.158 (-0.58)	-0.358 (-2.25)
Vocational	-0.076 (-1.32)	-0.228 (-0.91)	0.049 (0.14)	0.130 (0.62)	-0.244 (-0.47)	0.054 (0.14)
Other	0.019 (0.70)	0.065 (0.50)	-0.612 (-3.47)	-0.155 (-1.45)	0.017 (0.04)	-0.044 (-0.27)
R Square	0.493	0.310	0.346	0.261	0.289	0.122

Table 6. Estimated content factor correlation

NAEP92' Grade12

Num & Op	1.000				
Measurement	1.000	1.000			
Geometry	0.983	0.983	1.000		
Data Analysis	0.969	0.969	0.948	1.000	
Algebra	0.990	0.980	0.976	0.953	1.000

NAEP92' Grade8

Num & Op	1.000				
Measurement	.879	1.000			
Geometry	.945	.844	1.000		
Data Analysis	.985	.878	.943	1.000	
Algebra	.985	.877	.943	.982	1.000

Table 8. Standardized coefficients (& t-values) from the structural model (NAEP '92 grade 8)

	General	M-Meas	M-Geom	M-Data	C-Number	C-Geom
Female	0.048 (2.31)	-0.466 -(6.28)	-0.258 -(3.13)	-0.130 -(1.24)	0.292 (2.23)	-0.014 -(0.22)
Ethnicity						
Black	-0.851 -(24.64)	-0.404 -(3.47)	-0.022 -(0.16)	-0.442 -(2.67)	0.415 (2.01)	-0.401 -(3.84)
Hispanic	-0.525 -(15.76)	-0.127 -(1.07)	-0.087 -(0.66)	-0.465 -(2.79)	-0.289 -(1.39)	-0.130 -(1.23)
Asian	0.229 (3.71)	-0.405 -(1.84)	-0.223 -(0.90)	-0.681 -(2.17)	-0.423 -(1.09)	-0.258 -(1.31)
Parent's Ed.	0.194 (16.61)	-0.015 -(0.37)	-0.017 -(0.38)	0.013 (0.22)	-0.154 -(2.11)	-0.026 -(0.70)
TOC						
Rural	-0.022 -(0.60)	-0.033 -(0.24)	0.041 (0.27)	0.037 (0.20)	0.426 (1.71)	-0.030 -(0.23)
Disadv-Urban	-0.287 -(7.81)	-0.307 -(2.32)	0.213 (1.46)	-0.213 -(1.14)	0.010 (0.05)	-0.033 -(0.27)
Adv-Urban	0.304 (8.44)	-0.117 -(0.90)	0.202 (1.41)	-0.155 -(0.85)	-0.449 -(1.96)	-0.231 -(1.99)
School-Type						
Catholic	0.129 (4.01)	-0.252 -(2.20)	-0.102 -(0.79)	0.066 (0.40)	0.258 (1.27)	-0.039 -(0.38)
Private	0.080 (1.99)	0.105 (0.72)	0.058 (0.35)	0.160 (0.76)	0.131 (0.50)	0.098 (0.76)
Algebra	0.548 (22.69)	-0.103 -(1.24)	-0.167 -(1.82)	0.159 (1.36)	-0.252 -(1.73)	-0.188 -(2.55)
R Square	0.381	0.102	0.035	0.084	0.166	0.035

BEST COPY AVAILABLE